Minimax Goodness-of-Fit Testing in Multivariate Nonparametric Regression

Yuri I. Ingster *
Department of Mathematics II,
St. Petersburg State Electrotechnical University,
Russia

and

Theofanis Sapatinas
Department of Mathematics and Statistics,
University of Cyprus,
Cyprus

Abstract

We consider an unknown response function f defined on $\Delta = [0,1]^d$, $1 \leq d \leq \infty$, taken at n random uniform design points and observed with Gaussian noise of known variance. Given a positive sequence $r_n \to 0$ as $n \to \infty$ and a known function $f_0 \in L_2(\Delta)$, we propose, under general conditions, a unified framework for the goodness-of-fit testing problem for testing the null hypothesis $H_0 : f = f_0$ against the alternative $H_1 : f \in \mathcal{F}$, $\|f - f_0\| \geq r_n$, where $\mathcal{F}$ is an ellipsoid in the Hilbert space $L_2(\Delta)$ with respect to the tensor product Fourier basis and $\|\cdot\|$ is the norm in $L_2(\Delta)$. We obtain both rate and sharp asymptotics for the error probabilities in the minimax setup. The derived tests are inherently non-adaptive.

Several illustrative examples are presented. In particular, we consider functions belonging to ellipsoids arising from the well-known multidimensional Sobolev and tensor product Sobolev norms as well as from the less-known Sloan-Woźniakowski norm and a norm constructed from multivariable analytic functions on the complex strip.

Some extensions of the suggested minimax goodness-of-fit testing methodology, covering the cases of general design schemes with a known product probability density function, unknown variance, other basis functions and adaptivity of the suggested tests, are also briefly discussed.

Keywords: Goodness-of-Fit Tests, Hypotheses Testing, Minimax Testing, Nonparametric Alternatives, Nonparametric Regression, Random Design.

AMS Subject Classification: 62G08, 62G10, 62G20

*Research was partially supported by RFBI Grant 08-01-00692-a and by Grant NSh–638.2008.1

1 Introduction

We consider the multivariate nonparametric regression model with a random uniform design. More precisely, we observe

$$x_i = f(t_i) + \xi_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where t_i are random design points, $t_i \in \Delta = [0,1]^d$, $1 \leq d \leq \infty$. In particular, we assume that $t_i = \{t_i^k\}$ are (for $k = 1, \ldots, d$ and $i = 1, \ldots, n$) independent and identically distributed (*iid*) random variables with a uniform distribution, i.e., $t_i^k \stackrel{iid}{\sim} \mathcal{U}(0,1)$. Moreover, we assume that, conditionally on $T_n = \{t_1, \ldots, t_n\}$, ξ_i are *iid* Gaussian random variables with mean zero and variance τ^2, i.e., $\xi_i \stackrel{iid}{\sim} \mathcal{N}(0, \tau^2)$, where τ^2 is assumed to be *known* with $0 < \tau^2 < \infty$.

Given a positive sequence $r_n \to 0$ as $n \to \infty$ and a *known* function $f_0 \in L_2(\Delta)$, where $L_2(\Delta)$ is the set of squared-integrable functions on Δ, we propose, under general conditions, a unified framework for the goodness-of-fit testing problem for testing the null hypothesis

$$H_0 : f = f_0 \tag{1.2}$$

against the alternative

$$H_1 : f \in \mathcal{F}, \, \|f - f_0\| \geq r_n, \tag{1.3}$$

where $\mathcal{F}$ is an ellipsoid in the Hilbert space $L_2(\Delta)$ with respect to the tensor product Fourier basis and $\|\cdot\|$ is the norm in $L_2(\Delta)$. (The set $\mathcal{F}$ corresponds to a "regularity constraint" on the response function f.)

We are interested in both rate and sharp asymptotics for the error probabilities in the minimax setup, i.e., we try to find the maximal rate of convergence of $r_n \to 0$ as $n \to \infty$ which provide nontrivial minimax testing, when certain constraints are imposed on the regularity of the response function f.

Although there is a plethora of research work in the literature on the estimation problem for response functions $f \in \mathcal{F}$ in (both univariate and multivariate) nonparametric regression (under various design schemes), much less attention has been paid to the hypotheses testing problem in this model, especially in the multivariate case. This work is devoted to the goodness-of-fit testing problem (1.2)–(1.3) in the nonparametric regression model (1.1).

Nonparametric goodness-of-fit testing was studied intensively during the last twenty years or so; however, main results were obtained for the detection of the response function $f \in L_2(\Delta)$, with $d = 1$, in the 1-variable Gaussian white noise model, i.e.,

$$dX(t) = f(t)dt + \varepsilon dW(t), \quad t \in [0,1], \tag{1.4}$$

where $W(t)$ is the standard Wiener process, with the noise level $\varepsilon \to 0$. In particular, rate and sharp asymptotics for the error probabilities in the minimax setup were obtained for various classes $\mathcal{F}$ of nonparametric alternatives. Moreover, under periodicity, the sharp asymptotics are of Gaussian type and are determined by a specific extremal problem (see, e.g., [7], [8], [14], [18]).

These results have been extended in part to the density, spectral density, nonparametric regression and Poisson models for the 1-variable case (see, e.g., [8], [14], [17],

[18]). Note that, under some regularity constraints, one can formally deduce some results for the 1-variable density and nonparametric regression models from results on the asymptotic equivalence (in Le Cam sense) of these models to the 1-variable Gaussian white noise model (see, e.g., [2], [26]).

For the d-variable Gaussian white noise model, we have typically similar separation rates with the smoothness parameter σ (associated with the "regularity constraint" on the response function f) replaced by $\tilde{\sigma} = \sigma/d$ as well as sharp asymptotics of a similar type (see [19]). This leads to the "curse of dimensionality" phenomenon when d is large (see [20]). It was recently shown that one can actually lift the curse of dimensionality by using different type of regularity constraints, which are determined by the so-called "Sloan-Woźniakowski" norm (see [20]). Although, analogously to the 1-variable case, one can formally deduce, under some stronger regularity constraints, some results for the multivariate nonparametric regression models from results on the asymptotic equivalence (in Le Cam sense) of these models to the d-variable Gaussian white noise model (see, e.g., [3], [27]), one cannot apply these results to the tensor product Sobolev or Sloan-Woźniakowski type spaces, because there are no asymptotic equivalence results as yet for these spaces.

Rate asymptotics in d-variable parametric regression models were studied in, e.g., [9], [11], for testing a parametric model against Lipschitz and Hölder classes $\mathcal{F}$ of alternatives, respectively. On the other hand, rate asymptotics in the multivariate regression model, under equispaced design points, were studied in [1] for the goodness-of-fit testing problem (1.2)–(1.3), under Besov balls $\mathcal{F}$ of alternatives.

The purpose of this paper is to extend some results on the goodness-of-fit testing of [7], [14], [18]-[21] for the d-variable Gaussian white noise model to the goodness-of-fit testing problem (1.2)–(1.3) for the multivariate nonparametric regression model (1.1), in a unified framework.

In our study, we use analytic results on an extermal problem for ellipsoids that were presented in [14], [18]-[21] for the d-variable Gaussian white noise model. These lead to the asymptotic efficiency of testing for the multivariate nonparametric regression model (1.1), similar to the ones that have earlier been obtained, in specific settings, for the d-variable Gaussian white noise model, under the standard calibration $\varepsilon = \tau/\sqrt{n}$. However, the machinery of reduction of the hypothesis testing problems to the extermal problem is different and, essentially, more difficult, especially for the study of the lower bounds. The proposed tests are of different structure as well: they are based on U-statistics of increasing dimension. Certainly, this reduction requires some assumptions on the basis functions and on the sample size (compare with [6] for estimation problem). It is a typical situation for extending results from the Gaussian white noise model to other statistical models (e.g., density, spectral density, intensity of a Poisson process and so on).

Several illustrative examples are presented. In particular, we consider functions belonging to the balls under the well-known multidimensional Sobolev and tensor product Sobolev norms as well as from the less-known Sloan-Woźniakowski norm and a norm constructed from multivariable analytic functions on the complex strip. Some extensions of the suggested minimax goodness-of-fit testing methodology, covering the cases of general design schemes with a known product probability density function, unknown variance, other basis functions and adaptivity of the suggested tests, are also briefly

discussed.

2 Preliminaries and assumptions

2.1 Minimax goodness-of-fit testing

Consider the multivariate nonparametric regression model (1.1). Given a known function $f_0 \in L_2(\Delta)$, we test the null hypothesis (1.2), i.e., we test $H_0 : f = f_0$. Given a positive sequence $r_n \to 0$ as $n \to \infty$, let

$$\mathcal{F}(r_n) = \{f \in \mathcal{F} : \|f - f_0\| \geq r_n\},$$

where $\mathcal{F}$ is an ellipsoid in the Hilbert space $L_2(\Delta)$ with respect to the tensor product Fourier basis and $\|\cdot\|$ is the norm in $L_2(\Delta)$. Consider now the alternative hypothesis (1.3), i.e., consider $H_1 : f \in \mathcal{F}(r_n)$. (In what follows, without loss of generality, we restrict ourselves to the cases $f_0 = 0$ and $\tau = 1$.)

Set $X_n = \{x_1, \ldots, x_n\}$ and recall that $T_n = \{t_1, \ldots, t_n\}$. Let $P_{n,f}$ be the probability measure that corresponds to $Z_n = (X_n, T_n)$ and denote by $E_{n,f}$ the expectation over this probability measure. Let ψ be a (randomized) test, i.e., a measurable function of the observation Z_n taking values in $[0, 1]$: the null hypothesis is rejected with probability $\psi(Z_n)$ and is accepted with probability $1 - \psi(Z_n)$. Let

$$\alpha(\psi) = E_{n,0}\psi$$

be its type I error probability, and let

$$\beta(\mathcal{F}, r_n, \psi) = \sup_{f \in \mathcal{F}(r_n)} E_{n,f}(1 - \psi)$$

be its maximal type II error probability. We consider two criteria of asymptotic optimality:

[1] The first one corresponds to the classical Neyman-Pearson criterion. For $\alpha \in (0, 1)$ we set

$$\beta(\mathcal{F}, r_n, \alpha) = \inf_{\psi :\, a(\psi) \leq \alpha} \beta(\mathcal{F}, r_n, \psi).$$

We call a sequence of tests $\psi_{n,\alpha}$ *asymptotically minimax* if

$$\alpha(\psi_{n,\alpha}) \leq \alpha + o(1), \quad \beta(\mathcal{F}, r_n, \psi_{n,\alpha}) = \beta(\mathcal{F}, r_n, \alpha) + o(1),$$

where $o(1)$ is a sequence tending to zero; here, and in what follows, unless otherwise stated, all limits are taken as $n \to \infty$.

[2] The second one corresponds to the total error probabilities. Let $\gamma(\mathcal{F}, r_n, \psi)$ be the sum of the type I and the maximal type II error probabilities, and let $\gamma(\mathcal{F}, r_n)$ be the minimax total error probability, i.e.,

$$\gamma(\mathcal{F}, r_n) = \inf_{\psi} \gamma(\mathcal{F}, r_n, \psi),$$

where the infimum is taken over all possible tests. We call a sequence of tests ψ_n *asymptotically minimax* if

$$\gamma(\mathcal{F}, r_n, \psi_n) = \gamma(\mathcal{F}, r_n) + o(1).$$

It is known that (see, e.g., Chapter 2 of [18]) that

$$\beta(\mathcal{F}, r_n, \alpha) \in [0, 1-\alpha], \quad \gamma(\mathcal{F}, r_n) = \inf_{\alpha \in (0,1)} (\alpha + \beta(\mathcal{F}, r_n, \alpha)) \in [0, 1].$$

We consider the problems of rate and sharp asymptotics for the error probabilities in the minimax setup. The rate optimality problem corresponds to the study of the conditions for which $\gamma(\mathcal{F}, r_n) \to 1$ and $\gamma(\mathcal{F}, r_n) \to 0$ and, under the conditions of the last relation, to the construction of asymptotically *minimax consistent* sequences ψ_n, i.e, such that $\gamma(\mathcal{F}, r_n, \psi_n) \to 0$. Often, these conditions correspond to some minimal decreasing rates for the sequence r_n. Namely, we say that the positive sequence $r_n^* = r_n^*(\mathcal{F})$, $r_n^* \to 0$, is a *separation rate*, if

$$\gamma(\mathcal{F}, r_n) \to 1 \quad \text{as} \quad r_n/r_n^* \to 0,$$

and

$$\gamma(\mathcal{F}, r_n) \to 0, \quad \text{and} \ \ \beta(\mathcal{F}, r_n, \alpha) \to 0 \ \ \text{for any } \alpha \in (0,1), \quad \text{as} \quad r_n/r_n^* \to \infty.$$

In other words, it means that, for large n, one can detect all functions in $f \in \mathcal{F}$ if the ratio r_n/r_n^* is large, whereas if this ratio is small then it is impossible to distinguish between the null and the alternative hypothesis, with small minimax total error probability. Hence, the rate optimality problem corresponds to finding the separation rates r_n^* and to constructing asymptotically minimax consistent sequence of tests.

On the other hand, the sharp optimality problem corresponds to the study of the asymptotics of the quantities $\beta(\mathcal{F}, r_n, \alpha)$, $\gamma(\mathcal{F}, r_n)$ (up to vanishing terms) and to the construction of asymptotically minimax sequences $\psi_{n,\alpha}$, ψ_n, respectively. Often, the sharp asymptotics are of Gaussian type, i.e.,

$$\beta(\mathcal{F}, r_n, \alpha) = \Phi(H^{(\alpha)} - u_n) + o(1), \quad \gamma(\mathcal{F}, r_n) = 2\Phi(-u_n) + o(1), \tag{2.1}$$

where Φ is the standard Gaussian distribution function, $H^{(\alpha)}$ is its $(1-\alpha)$-quantile, i.e., $\Phi(H^{(\alpha)}) = 1-\alpha$, and the sequence $u_n = u_n(\mathcal{F}, r_n)$ characterizes *distinguishability* in the problem. The separation rates r_n^* are usually determined by the relation $u_n(\mathcal{F}, r_n^*) \asymp 1$ (see, e.g., [14], [18]). Hence, the sharp optimality problem corresponds to calculating the sequence u_n and to constructing asymptotically minimax sequence of tests.

2.2 Assumptions

Let $L_2(\Delta) = L_2$, $\mathcal{L}$ be a denumerable set, $\{\phi_l\}_{l \in \mathcal{L}}$ be an orthonormal system in L_2, and $L_2^{\mathcal{L}} \subset L_2$ be the closed linear hull of the system $\{\phi_l\}_{l \in \mathcal{L}}$. For a function $f \in L_2^{\mathcal{L}}$, let $\theta = \{\theta_l\}_{l \in \mathcal{L}}$ be the "generalized" Fourier coefficients with respect to this system, i.e., $\theta_l = \langle f, \phi_l \rangle$, $l \in \mathcal{L}$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in L_2.

Let a collection of coefficients $\{c_l\}_{l \in \mathcal{L}}$, $c_l \geq 0$, be given. The set of functions $\mathcal{F} \subset L_2^{\mathcal{L}}$ under consideration are the *ellipsoids* with respect to the orthonormal system $\{\phi_l\}_{l \in \mathcal{L}}$ with coefficients $\{c_l\}_{l \in \mathcal{L}}$, $l \in \mathcal{L}$, i.e.,

$$\mathcal{F} = \{f \ : \ f(t) = \sum_{l \in \mathcal{L}} \theta_l \phi_l(t), \ \sum_{l \in \mathcal{L}} c_l^2 \theta_l^2 \leq 1\}.$$

Let
$$\mathcal{N}(C) = \{l \in \mathcal{L} : c_l < C\}, \qquad N(C) = \#\mathcal{N}(C),$$
where # denotes the cardinality of a set.

Consider the following set of assumptions:

(A1) The set $\mathcal{N}(C)$ is finite, i.e.,
$$N(C) < \infty \quad \forall\, C > 0.$$

(A2) The orthonormal system $\{\phi_l\}_{l\in\mathcal{L}}$ satisfies
$$\sum_{l\in\mathcal{N}(C)} \phi_l^2(t) = N(C) \quad \forall\, C > 0,\ t \in \Delta.$$

(A3) The functions $f \in \mathcal{F}$ are uniformly bounded in $L_p(\Delta)$-norm for some $p > 4$, i.e.,
$$\exists\, p > 4: \quad \sup_{f\in\mathcal{F}} \int_\Delta |f(t)|^p < \infty.$$

Remark 2.1 Note that assumption **(A3)** follows from the following stronger condition,
$$\sup_{f\in\mathcal{F}} \|f\|_\infty < \infty, \tag{2.2}$$
where $\|f\|_\infty = \sup_{t\in\Delta} |f(t)|$.

3 Rate optimality

In what follows, the relation $A_n \sim B_n$ means that A_n/B_n tends to 1 while the relation $A_n \asymp B_n$ means that there exists constants $0 < c_1 \le c_2 < \infty$ and n_0 large enough such that $c_1 \le A_n/B_n \le c_2$ for $n \ge n_0$. Let also $\mathbb{1}_{\{A\}}$ be the indicator function of a set A.

For a sequence $C = C_n$, let $\mathcal{N} = \mathcal{N}(C_n)$, $N = N(C_n)$.

Let us introduce an extra assumption.

(B1) $\qquad N = o(n).$

Theorem 1 *Let $r_n \to 0$.*

(1)[Lower bounds] Assume **(A1)**–**(A2)***. Take $C_n \to \infty$ such that $\limsup(C_n r_n) < 1$ and* **(B1)** *holds. Then*
$$\beta(\mathcal{F}, r_n, \alpha) \ge \Phi(H^{(\alpha)} - u_n) + o(1), \quad \gamma(\mathcal{F}, r_n) \ge 2\Phi(-u_n) + o(1),$$
where
$$u_n^2 = \frac{n^2 r_n^4}{2N}. \tag{3.1}$$

(2) [Upper bounds] Assume **(A1)**–**(A3)***. Take $C_n \to \infty$ such that* **(B1)** *holds. Consider the sequence of tests $\psi_n^H = \mathbb{1}_{\{U_n > H\}}$ based on the U-statistics*
$$U_n = \frac{1}{n} \sum_{1\le i<k\le n} K_n(z_i, z_k), \tag{3.2}$$

where $z_i = (x_i, t_i),\ i = 1, \dots, n$ are the observations, with the kernel

$$K_n(z', z'') = x' x'' G_n(t', t''), \quad G_n(t', t'') = \sqrt{\frac{2}{N}} \sum_{l \in \mathcal{N}} \phi_l(t')\phi_l(t''). \tag{3.3}$$

Set

$$h_n(f) = \frac{n}{\sqrt{2N}} \sum_{l \in \mathcal{N}} \theta_l^2. \tag{3.4}$$

Then, uniformly over $H = H_n \in \mathbb{R}$,

$$\alpha(\psi_n^H) \le 1 - \Phi(H) + o(1),$$

and, for any $c \in (0,1)$, uniformly over $f \in \mathcal{F}$ and $H = H_n$ such that $h_n(f) \ge cH_n$,

$$\beta(\mathcal{F}, r_n, \psi_n^H) \le \Phi(H - h_n(f)) + o(1).$$

Remark 3.1 We now give some intuition about the suggested U-statistics used in Theorem 1. For testing the null hypothesis $H_0 : f = 0$ in the Gaussian white noise model, a natural test statistic is a centered and normalized (under H_0) version of the quadratic functional $\sum_{l \in \mathcal{L}} \hat{\theta}_l^2$, where $\hat{\theta}_l = \int_\Delta \phi_l(t) dX(t)$. The analog of $\hat{\theta}_l$ in the multivariate nonparametric regression model (1.1) is given by $\hat{\theta}_l = n^{-1} \sum_{i=1}^n \phi_l(t_i) x_i$ which leads to the quadratic functional

$$\sum_{l \in \mathcal{L}} \hat{\theta}_l^2 = \frac{1}{n^2} \sum_{i,k=1}^n x_i x_k \widetilde{G}_n(t_i, t_k), \quad \widetilde{G}_n(t', t'') = \sum_{l \in \mathcal{L}} \phi_l(t')\phi_l(t'').$$

Suppressing now the terms with $i = k$, a centered and normalized version of this quadratic functional corresponds to the U-statistic defined in (3.2) with the kernel defined in (3.3).

Let the sequence $C = C_n$ be determined by the "balance equation"

$$C_n^4 N(C_n) \asymp n^2. \tag{3.5}$$

Observe that, in this case, under **(A1)**, $C_n \to \infty$ and, hence, $N(C_n) \to \infty$.

Remark 3.2 Note that if r_n satisfies $C_n r_n \asymp 1$, then (3.5) corresponds to $u_n \asymp 1$ in (3.1). Corollaries 1 and 2 below show a motivation of (3.5).

Let us introduce an extra assumption.

(B2) For any $B > 0$, $N(C_n) \asymp N(BC_n)$.

Note that we can obtain lower bounds for $h_n(f)$ from (3.4). Indeed, for $f \in \mathcal{F}(r_n)$, we have

$$\begin{aligned} h_n(f) &= \frac{n}{\sqrt{2N}} \left(\sum_{l \in \mathcal{L}} \theta_l^2 - \sum_{c_l \ge C_n} \theta_l^2 \right) \ge \frac{n}{\sqrt{2N}} \left(r_n^2 - C_n^{-2} \sum_{c_l \ge C_n} c_l^2 \theta_l^2 \right) \\ &\ge \frac{n}{\sqrt{2N}} (r_n^2 - C_n^{-2}) = \frac{n r_n^2}{\sqrt{2N}} \left(1 - (r_n C_n)^{-2}\right). \end{aligned} \tag{3.6}$$

Therefore, if $C_n r_n \geq B > 1$, we have from Theorem 1 (2),

$$\beta(\mathcal{F}, r_n, \psi_n^H) \leq \Phi\big(H - u_n(1 - B^{-2})\big) + o(1),$$

with u_n determined by (3.1). This leads to

Corollary 1 *Let $r_n \to 0$. Assume* **(A1)**–**(A3)** *and* **(B1)**–**(B2)**. *Then*

[1] The separation rates are of the form

$$r_n^* \asymp C_n^{-1},$$

where the sequence $C = C_n$ is determined by (3.5).

[2] Moreover, let $r_n/r_n^* \to \infty$. Then, there exists a sequence $H = H_n \to \infty$ such that the sequence of tests $\psi_n^H = \mathbb{1}_{\{U_n > H\}}$ is asymptotically minimax consistent, i.e., $\gamma(\mathcal{F}, r_n, \psi_n^H) \to 0$.

We say that a function $g(t),\ t > 0$, is a *slowly varying* function if $g(Bt)/g(t)$ tends to 1 as $t \to \infty$, for any $B > 0$.

This leads to the following assumption.

(B3) $N(C_n)$ is a slowly varying function.

Corollary 2 *Let $r_n \to 0$. Assume* **(A1)**–**(A3)** *and* **(B1)**–**(B3)**. *Then*

[1] The sharp asymptotics (2.1) hold, where u_n is defined by (3.1) with any $N(C_n)$ determined by (3.5).

[2] Moreover, for any sequence C_n satisfying (3.5), there exists a sequence $B_n \to \infty$ such that, for the sequence $C_{n,1} = B_n C_n$, the sequence of tests $\psi_n^{H^{(\alpha)}}$ is asymptotically minimax under the Neyman-Pearson criterion, and the sequence of tests $\psi_n^{u_n/2}$ is asymptotically minimax under the total error probability criterion.

Proof. In order to get the upper bounds, note that under **(B3)** one can take a sequence $B_n \to \infty$ such that $N(B_n C_n) \sim N(C_n)$. Applying Theorem 1 (2) for the sequence $C_{n,1} = B_n C_n$, and for $H = H^{(\alpha)}$ and $H = u_n/2$, and recalling (3.6), we obtain

$$\inf_{f \in \mathcal{F}(r_n)} h_n(f) \geq u_n(1 + o(1)).$$

By (3.4), Corollary 2 (2) now follows.

In order to get the lower bounds, observe first that asymptotics of u_n do not depend on a sequence C_n involved in (3.5). In fact, if $C_{n,0}$ is another sequence applicable to (3.5), then $C_{n,0} \sim B_n C_n,\ B_n \asymp 1$ and, under **(B3)**, we have $N(C_{n,0}) \sim N(C_n)$. Fix now a sequence C_n in (3.5). It suffices to consider the case $u_n \asymp 1$, which corresponds to having $r_n C_n \sim A_n \asymp 1$. By taking another sequence $C_{n,0} = B_n C_n,\ B_n \sim (2A_n)^{-1}$, we get $r_n C_{n,0} \sim 1/2$. Applying Theorem 1 (1), Corollary 2 (1) now follows. This completes the proof of Corollary 2. $\square$

4 Sharp optimality

4.1 Extremal problem

In order to describe the sharp asymptotics similar to [14], [18], we have to consider an extremal problem on the space of collections $v = \{v_l\}_{l\in\mathcal{L}}$.

Assume that $r_n \to 0$. For $b = b_n \asymp 1, B = B_n \asymp 1$, by following arguments similar to those in Chapter 4 of [18], we arrive at

$$u_n^2(b,B) = \inf_{v\in V_n(b,B)} \frac{1}{2}\sum_{l\in\mathcal{L}} v_l^4, \tag{4.1}$$

$$V_n(b,B) = \left\{v : \sum_{l\in\mathcal{L}} v_l^2 \geq n(Br_n)^2, \quad \sum_{l\in\mathcal{L}} c_l^2 v_l^2 \leq nb^2\right\}. \tag{4.2}$$

Let $u_n(B) = u_n(1,B)$ and $u_n = u_n(1,1)$. From Proposition 2.8 of [18], it follows that $u_n^2(b,B)$ is a convex function in (b^2, B^2) and, from rescaling arguments, it is easily seen that $u_n^2(b,B) = b^4 u_n^2(B/b)$.

By using Lagrange multipliers, the extremal collection $v_n = \{v_{l,n}\}_{l\in\mathcal{L}}$ in (4.1) is of the form $v_{l,n}^2 = z_0^2(1-(c_l/C)^2)_+$, where $a_+ = \max(0,a)$ for any real number a, and the quantities $z_0 = z_{n,0}(b,B) > 0$, $C = C_n(b,B)$ are determined by the equations

$$\sum_{l\in\mathcal{L}} v_{l,n}^2 = z_0^2 \sum_{c_l<C}(1-(c_l/C)^2) = n(Br_n)^2, \tag{4.3}$$

$$\sum_{l\in\mathcal{L}} c_l^2 v_{l,n}^2 = z_0^2 \sum_{c_l<C} c_l^2(1-(c_l/C)^2) = nb^2, \tag{4.4}$$

while the value of the extremal problem is

$$u_n^2(b,B) = \frac{1}{2}\sum_{l\in\mathcal{L}} v_{l,n}^4 = \frac{1}{2}z_0^4 \sum_{c_l<C}(1-(c_l/C)^2)^2. \tag{4.5}$$

Let

$$\begin{aligned} I_1 &= \sum_{l\in\mathcal{N}}(1-(c_l/C)^2), \quad I_0 = \sum_{l\in\mathcal{N}}(1-(c_l/C)^2)^2, \\ I_2 &= \sum_{l\in\mathcal{N}}(c_l/C)^2(1-(c_l/C)^2). \end{aligned}$$

It is easily seen that the equations (4.3)–(4.5) can be rewritten in the form

$$z_0^2 I_1 = n(Br_n)^2, \quad C^2 z_0^2 I_2 = nb^2, \quad u_n^2(b,B) = \frac{1}{2}z_0^4 I_0 = \frac{n^2(Br_n)^4 I_0}{2I_1^2}. \tag{4.6}$$

Observe that $I_1 = I_0 + I_2 \geq I_2$ and

$$C^2 = \frac{b^2 I_1}{I_2 B^2 r_n^2} \geq b^2(Br_n)^{-2} \to \infty \quad \text{as} \quad r_n \to 0.$$

Under (**A1**), this yields $N \to \infty$. Moreover, one has

$$(3/4)N(C/2) \leq I_1 \leq N(C), \quad (3/4)^2 N(C/2) \leq I_0 \leq N(C).$$

Hence, under (**B2**), these yield

$$I_1 \asymp I_0 \asymp N, \quad z_0^2 \asymp \frac{nr_n^2}{N}, \quad u_n^2(b,B) \asymp \frac{n^2 r_n^4}{N}. \tag{4.7}$$

Introduce the additional assumption

(**C1**) For all $B = B_n \asymp 1$, $u_n(B) \asymp u_n$.

Note that, under assumption (**C1**), we get

$$u_n^2(b,B) \sim u_n^2 \quad \text{as} \quad b = b_n \to 1,\ B = B_n \to 1.$$

(compare with Propositions 2.8 and 5.6 in [18]).

4.2 Sharp asymptotics

Theorem 2 *Let $r_n \to 0$.*

*(1) [Lower bounds] Assume (**A1**)–(**A2**), (**B1**)–(**B2**) and (**C1**). Then*

$$\beta(\mathcal{F}, r_n, \alpha) \geq \Phi(H^{(\alpha)} - u_n) + o(1), \quad \gamma(\mathcal{F}, r_n) \geq 2\Phi(-u_n/2) + o(1), \tag{4.8}$$

where u_n is the value of the extremal problem (4.1), (4.2) for $b = B = 1$.

*(2) [Upper bounds] Assume (**A1**)–(**A3**) and (**B1**)–(**B2**). Let $\liminf u_n > 0$. Consider the sequence of tests $\psi_n^H = 1\!\!I_{\{U_n > H\}}$ based on the U-statistics*

$$U_n = \frac{1}{n} \sum_{1 \leq i < k \leq n} K_n(z_i, z_k),$$

where $z_i = (x_i, t_i),\ i = 1, \dots, n$, are the observations, with the kernel

$$K_n(z', z'') = x' x'' G_n(t', t''), \quad G_n(t', t'') = \sum_{l \in \mathcal{N}} w_{n,l} \phi_l(t') \phi_l(t''), \tag{4.9}$$

where $w_{n,l} = v_{l,n}^2 / u_n$ and $\{v_{l,n}\}$ is the extremal sequence of the extremal problem (4.1), (4.2) for $b = B = 1$, or, equivalently,

$$w_{n,l} = (1 - (c_l/C)^2)_+ / w_n, \quad w_n^2 = \frac{1}{2} \sum_{l \in \mathcal{N}} (1 - (c_l/C)^2)^2.$$

Then, uniformly over $H = H_n \in \mathbb{R}$,

$$\alpha(\psi_n^H) \leq 1 - \Phi(H) + o(1),$$

and, for any $c \in (0,1)$, uniformly over $H = H_n$ such that $u_n \geq cH_n$,

$$\beta(\mathcal{F}, r_n, \psi_n^H) \leq \Phi(H - u_n) + o(1). \tag{4.10}$$

Remark 4.1 Combining (4.8) and (4.10), we see that the sequence of tests ψ_n^H with $H = H^{(\alpha)}$ is asymptotically minimax under the Neyman-Pearson criterion, i.e.,

$$\alpha(\psi_n^{H^{(\alpha)}}) \leq \alpha + o(1), \quad \beta(\mathcal{F}, r_n, \psi_n^{H^{(\alpha)}}) = \Phi(H^{(\alpha)} - u_n) + o(1),$$

and the sequence of tests ψ_n^H with $H = u_n/2$ is asymptotically minimax under the total error probability criterion, i.e.,

$$\gamma(\mathcal{F}, r_n, \psi_n^{u_n/2}) = 2\Phi(-u_n/2) + o(1).$$

5 Tensor product Fourier basis

Let $\mathbb{Z}_*^\infty \subset \mathbb{Z}^\infty$ consists of all sequences $l = (l_1, \ldots, l_d, \ldots)$ with finite number j such that $l_j \neq 0$, and consider the natural embedding $\mathbb{Z}^d \subset \mathbb{Z}_*^\infty : (l_1, \ldots, l_d) \to (l_1, \ldots, l_d, 0, \ldots)$. Let $\mathcal{L}$ be an infinite subset of $\mathbb{Z}_*^\infty$.

Consider the tensor product Fourier basis $\{\phi_l\}_{l\in\mathcal{L}}$ in L_2, i.e.,

$$\phi_l(t) = \prod_k \phi_{l_k}(t^k), \quad t = (t^1, \ldots t^d, \ldots) \in \Delta, \quad l \in \mathcal{L}, \tag{5.1}$$

where $\phi_j(u),\ j \in \mathbb{Z},\ u \in [0, 1]$, is the standard Fourier basis in $L_2([0, 1])$, i.e.,

$$\phi_0(u) = 1, \quad \phi_j(u) = \sqrt{2}\cos(2\pi ju), \quad \phi_{-j}(u) = \sqrt{2}\sin(2\pi ju), \quad j > 0.$$

Definition 5.1 *A set $\mathcal{L}$ is called sign-symmetric if, for all $l = (l_1, \ldots, l_d, \ldots) \in \mathcal{L}$, one has $\varepsilon l = (\varepsilon_1 l_1, \ldots, \varepsilon_d l_d, \ldots) \in \mathcal{L}$ for all $\varepsilon_j = \pm 1$.*

Definition 5.2 *The collection $\{h_l\}_{l\in\mathcal{L}}$ is called sign-symmetric if the set $\mathcal{L}$ is sign-symmetric and $h_l = h_{\varepsilon l}$ for all $l \in \mathcal{L}$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_d, \ldots),\ \varepsilon_j = \pm 1$.*

(D1) The set $\mathcal{L}$ and the collection of coefficients $\{c_l\}_{l\in\mathcal{L}}$ are sign-symmetric.

Let us now show that, under assumptions **(A1)** and **(D1)**, assumption **(A2)** holds true for the tensor product Fourier basis (5.1). Since the set $\mathcal{N}$ is sign-symmetric then, under assumption **(D1)**, this follows from the following statement.

Lemma 5.1 *Let $\mathcal{M} \subset \mathbb{Z}_*^\infty$ be a finite sign-symmetric set and let $\{\phi_l\}_{l\in\mathcal{L}}$ be the tensor product Fourier basis (5.1). Then*

$$\sum_{l\in\mathcal{M}} \phi_l^2(t) = \#(\mathcal{M}) \quad \forall\, t \in \Delta.$$

Proof. Consider the presentation $\mathcal{M} = \cup_u \mathcal{M}_u$, where $u \subset \mathbb{N}$ and $\mathcal{M}_u$ consists of $l \in \mathcal{M}$ such that $\#\{j : l_j \neq 0\} = m$. It suffices to check that, for all u,

$$\sum_{l\in\mathcal{M}_u} \phi_l^2(t) = \#(\mathcal{M}_u) \quad \forall\ t \in \Delta.$$

Clearly, this holds for $u = \emptyset$. Without loss of generality, assume $m = \{1, \ldots, d\},\ d \in \mathbb{N}$. Let $\mathcal{M}_u^+ = \{l \in \mathcal{M}_u : l_j > 0\ \forall\ j \in u\}$. Since $\mathcal{M}$ is sign-symmetric, $\mathcal{M}_u^+$ consists of all $\bar{\varepsilon}l,\ l \in \mathcal{M}_u^+,\ \bar{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d),\ \varepsilon_k = \pm 1$ and $\#(\mathcal{M}_u) = 2^d \#(\mathcal{M}_u^+)$. It suffices then to check that, for each $l \in \mathcal{M}_u^+$,

$$\sum_{\bar{\varepsilon}} \phi_{\varepsilon l}^2(t) = 2^d.$$

Consider $\varepsilon_k,\ k = 1, \ldots, d$, as *iid* Rademacher random variables, i.e., $P(\varepsilon_k = 1) = P(\varepsilon_k = -1) = 1/2$. Then, by independency,

$$\sum_{\bar{\varepsilon}} \phi_{\varepsilon l}^2(t) = 2^d E_{\bar{\varepsilon}} \prod_{k=1}^d \phi_{\varepsilon_k l_k}^2(t^k) = 2^d \prod_{k=1}^d E_{\varepsilon_k} \phi_{\varepsilon_k l_k}^2(t^k) = 2^d,$$

since $E_{\varepsilon_k} \phi_{\varepsilon_k l_k}^2(t^k) = (2\sin^2(l_k t^k) + 2\cos^2(l_k t^k))/2 = 1$. This completes the proof of Lemma 5.1. $\Box$

Remark 5.1 Note that for the tensor product Fourier basis (5.1), condition (2.2) (and, hence, assumption **(A3)**) is fulfilled if

$$\sum_{l\in\mathcal{L}} 2^{J(l)} c_l^{-2} < \infty, \quad J(l) = \#\{j : l_j \neq 0\}. \tag{5.2}$$

Indeed, we have $\sup_{t\in\Delta} |\phi_l(t)| = 2^{J(l)/2}$, and hence

$$\begin{aligned} \|f\|_\infty^2 \le \left(\sum_{l\in\mathcal{L}} |\theta_l| \sup_{t\in\Delta} |\phi_l(t)|\right)^2 &\le \left(\sum_{l\in\mathcal{L}} \theta_l^2 c_l^2\right)\left(\sum_{l\in\mathcal{L}} 2^{J(l)} c_l^{-2}\right) \\ &\le \sum_{l\in\mathcal{L}} 2^{J(l)} c_l^{-2}. \end{aligned}$$

6 Examples: rate and sharp asymptotics in various ellipsoids

Let us first give some extra notation. For the function $f = \sum_{l\in\mathcal{L}} \theta_l \phi_l \in L_2^{\mathcal{L}}$, we set $\|f\|_c^2 = \sum_{l\in\mathcal{L}} \theta_l^2 c_l^2$ and let $L_{2,c}^{\mathcal{L}} = \{f \in L_2^{\mathcal{L}} : \|f\|_c < \infty\}$ be the Hilbert space with the norm $\|\cdot\|_c$. (Clearly the ellipsoid $\mathcal{F}$ is the unit ball in $L_{2,c}^{\mathcal{L}}$.)

Consider the tensor product Fourier basis (5.1). In all examples below, assumption **(D1)** holds true. Hence, by Lemma 5.1, assumption **(A2)** holds true. It is easily seen that assumption **(A1)** is also fulfilled in all examples below. That the assumption **(A3)** holds also true is discussed in each example separately.

The first two examples are versions of the classical multidimensional Sobolev norm (see [19]).

6.1 Multidimensional Sobolev norms

Let $\Delta = [0,1]^d,\ d \in \mathbb{N},\ \mathcal{L} = \mathbb{Z}^d \setminus \{0\}$, and let

$$c_l^2 = \sum_{k=1}^{d} |2\pi l_k|^{2\sigma},\ l \in \mathcal{L},\ \sigma > 0. \tag{6.1}$$

Then, for $\sigma \in \mathbb{N}$, the norm $\|f\|_c$ corresponds to the sum of σ-derivatives of a 1-periodic f over all variables, i.e.,

$$\|f\|_c^2 = \sum_{k=1}^{d} \|\partial^\sigma f / \partial t_k^\sigma\|^2, \tag{6.2}$$

where $\|\cdot\|$ is the norm in $L_2(\Delta)$.

Assumption **(A3)** is fulfilled for $\sigma > d/4$ by the so-called Sobolev embedding theorem (see Eq. (3.2.20) of [5]).

Let now

$$c_l^2 = \left(\sum_{k=1}^{d} (2\pi l_k)^2\right)^\sigma,\ l \in \mathcal{L},\ \sigma > 0. \tag{6.3}$$

Then, for $\sigma \in \mathbb{N}$, the norm $\|f\|_c$ corresponds to the sum of all the derivatives of a 1-periodic f of order σ, i.e.,

$$\|f\|_c^2 = \sum_{i_1=1}^{d} \cdots \sum_{i_\sigma=1}^{d} \|\partial^\sigma f/\partial t_{i_1} \ldots \partial t_{i_\sigma}\|^2. \tag{6.4}$$

Certainly, the norms (6.2) and (6.4) are equivalent for any fixed d since the ratio of coefficients in (6.1) and (6.3) is bounded and away from 0. Hence, assumption **(A3)** is fulfilled for $\sigma > d/4$.

It was shown in [19] that

$$N(C) \sim C^{d/\sigma} J_k(d,\sigma), \;\; k = 1, 2,$$

(e.g., $k = 1$ corresponds to (6.1) and (6.2), and $k = 2$ corresponds to (6.3) and (6.4)), where

$$J_1(d,\sigma) = \frac{\Gamma^d(1 + 1/2\sigma)}{\pi^d \Gamma(1 + d/2\sigma)}, \quad J_2(d,\sigma) = \frac{1}{2^d \pi^{d/2} \Gamma(1 + d/2)}.$$

Using equation (3.5), these yield

$$C \asymp n^{2\sigma/(4\sigma+d)}, \quad N(C) \asymp n^{2d/(4\sigma+d)}.$$

Hence, assumption (**B2**) is fulfilled while assumption (**B1**) is fulfilled for $\sigma > d/4$. Thus, we obtain the separation rates

$$r_n^* = n^{-2\sigma/(4\sigma+d)}.$$

For the sharp asymptotics, it was shown that

$$u_n^2 \sim C_k(d,\sigma) n^2 r_n^{4+d/\sigma}, \quad k = 1, 2,$$

where, for the norm (6.2),

$$C_1(d,\sigma) = \frac{\pi^d (1 + 2\sigma/d)\Gamma(1 + d/2\sigma)}{(1 + 4\sigma/d)^{1+d/2\sigma} \Gamma^d(1 + 1/2\sigma)},$$

and for the norm (6.4),

$$C_2(d,\sigma) = \frac{\pi^d (1 + 2\sigma/d)\Gamma(1 + d/2)}{(1 + 4\sigma/d)^{1+d/2\sigma} \Gamma^d(3/2)}.$$

Assumption (**C1**) is thus fulfilled. Hence, we arrive at (2.1).

The next two examples correspond to tensor product norms in ANOVA modeling. These spaces are capable of dealing with interactions of all orders in a flexible way, thus vastly extending the classical additive methodology in multivariate nonparametric regression inference (see [12], [25]).

6.2 Tensor product Sobolev norm

Let $\Delta = [0,1]^d,\ d \in \mathbb{N},\ \mathcal{L} = \mathbb{Z}^d$, and let

$$c_l = \prod_{k: l_k \neq 0} |2\pi l_k|^\sigma,\ l \in \mathcal{L},\ c_{0,\ldots,0} = 1. \tag{6.5}$$

For a $\sigma \in \mathbb{N}$, this corresponds to the following (see [25]). Let us consider the functional orthogonal ANOVA expansion

$$f(t) = \sum_u f_u(t_u), \quad \int_\Delta f_u(t_u) dt_k = 0 \quad \forall\ k \in u, \tag{6.6}$$

where the sum is taken over all subsets $u = \{j_1, \ldots j_m\} \subset \{1, \ldots, d\},\ 1 \le j_1 < \ldots < j_m \le d\}$ and $t_u = \{t_{j_1}, \ldots, t_{j_m}\}$, if $u = \emptyset$, then $f_u = \text{constant} = \int_\Delta f(t)dt$. Then,

$$\|f\|_c^2 = \sum_u \|f_u\|_{c,u}^2,$$

where $\|f_u\|_{c,u}$ is the norm of mixed $m\sigma$-derivatives of a 1-periodic f_u, i.e.,

$$\|f_u\|_{c,u} = \|\partial^{m\sigma} f / \partial t_{j_1}^\sigma \ldots \partial t_{j_m}^\sigma\|. \tag{6.7}$$

Assumption **(A3)** is fulfilled for $\sigma > 1/4$, using appropriate embedding properties (see Chapter III of [30]).

It was shown in [21] that

$$N(C) \sim \frac{C^{1/\sigma} \log^{d-1}(C)}{\pi^d \sigma^{d-1} \Gamma(d)}. \tag{6.8}$$

Using equation (3.5), this yields

$$C \asymp \left(\frac{n^2}{\log^{d-1}(n)} \right)^{\sigma/(4\sigma+1)}.$$

Hence, assumption (**B2**) is fulfilled while assumption (**B1**) is fulfilled for $\sigma > 1/4$. Thus, we obtain the separation rates

$$r_n^* = \left(\frac{\log^{d-1}(n)}{n^2} \right)^{\sigma/(4\sigma+1)}.$$

For the sharp asymptotics, it was shown that

$$u_n^2 \sim \frac{C(d,\sigma) n^2 r_n^{4+1/\sigma}}{\log^{d-1}(r_n^{-1})}, \tag{6.9}$$

where

$$C(d,\sigma) = \frac{2b(\sigma)\Gamma(d)(\pi\sigma)^d}{(1+4\sigma)^{b(\sigma)}}, \quad b(\sigma) = \frac{2\sigma+1}{2\sigma}. \tag{6.10}$$

Assumption (**C1**) is thus fulfilled. Hence, we arrive at (2.1).

6.3 ANOVA subspaces

Let $\Delta = [0,1]^d$, $d \in \mathbb{N}$. Taking $m \in \{0,1,\dots,d\}$, let $\mathcal{L}_m^d$ be the set that consists of $l \in \mathbb{Z}^d$ such that $\#\{k : l_k \neq 0\} = m$, and $\mathcal{L}^{d,m} = \bigoplus_{j=0}^m \mathcal{L}_j^d$. Under (6.6), the spaces $L_2^{\mathcal{L}_m^d}$ and $L_2^{\mathcal{L}^{d,m}}$ consist of the functions

$$f(t) = \sum_{u:\#(u)=m} f_u(t_u), \quad f(t) = \sum_{u:\#(u)\leq m} f_u(t_u),$$

respectively, i.e., they consist of sums of functions of m variables or no more than m variables. If $m = 0$, this corresponds to the constant function while the case $m = 1$ corresponds to functions with an additive structure. Take c_l according to (6.5). Then, we obtain,

$$\|f\|_c^2 = \sum_{u:\#(u)=m} \|f_u\|_{c,u}^2, \quad \|f\|_c^2 = \sum_{u:\#(u)\leq m} \|f_u\|_{c,u}^2,$$

respectively, where, for $\sigma \in \mathbb{N}$, the norm $\|f_u\|_{c,u}$ of a 1-periodic f_u is determined by (6.7) (see [25]). Assumption **(A3)** is fulfilled for $\sigma > 1/4$, since the spaces presented here are subspaces of the tensor product Sobolev spaces discussed in Section 6.3.

Take c_l according to (6.5). Denote by $N_d(C)$ the function $N(C)$ for the tensor product Sobolev norms, by $N_{d,m}(C)$ the function $N(C)$ for $\mathcal{L} = \mathcal{L}^{d,m}$, and by $N_m^d(C)$ the function $N(C)$ for $\mathcal{L} = \mathcal{L}_m^d$. Observe that

$$N_m^d(C) = \binom{d}{m} N_m^m(C), \quad N_{d,m}(C) = \sum_{j=0}^m \binom{d}{j} N_j^d(C).$$

Set $M = \binom{d}{m}$ and note that $M \geq 1$ for $0 \leq m \leq d$. It was shown in [21] that, as $C \to \infty$,

$$N_{d,m}(C) \sim M N_m^m(C) \sim M N_m(C) \sim \frac{M C^{1/\sigma} \log^{m-1}(C)}{\pi^m \sigma^{m-1} \Gamma(m)}, \tag{6.11}$$

the last relation follows from (6.8). For both the cases $\mathcal{L}_m^d$ and $\mathcal{L}^{d,m}$, using (3.5), we have

$$C \asymp \left(\frac{\tilde{n}^2}{\log^{m-1}(\tilde{n})} \right)^{\sigma/(4\sigma+1)}, \quad \tilde{n} \stackrel{\Delta}{=} n/\sqrt{M}.$$

Hence, assumption (**B2**) is fulfilled while assumption (**B1**) is fulfilled for $\sigma > 1/4$. Thus, we obtain the separation rates

$$r_n^* = \left(\frac{\log^{m-1}(\tilde{n})}{\tilde{n}^2} \right)^{\sigma/(4\sigma+1)}.$$

Let $u_{n,d}$ be the quantities that determine the sharp asymptotics for the tensor product Sobolev norms with sharp asymptotics (6.9). Using (6.11), we obtain, for both cases, the sharp asymptotics

$$u_n^2 \sim \frac{u_{n,m}^2}{M} \sim \frac{C(m,\sigma) n^2 r_n^{4+1/\sigma}}{M \log^{m-1}(r_n^{-1})}, \tag{6.12}$$

where the constant $C(m,\sigma)$ is defined by (6.10). (Note that (6.12) corresponds, in the case $m < d$, to some loss of efficiency compared to (6.9), since the sample size n is now

reduced by the factor $M^{-1/2} > 1$.) Assumption (**C1**) is thus fulfilled. Hence, we arrive at (2.1).

The next example corresponds to classical multivariable analytic functions on the complex strip (see [22], [24]).

6.4 Multivariable analytic functions on the complex strip

Let $\Delta = [0,1]^d,\ d \in \mathbb{N},\ \mathcal{L} = \mathbb{Z}^d$ and, for $\kappa > 0$, let

$$c_l^2 = \prod_{k=1}^{d} \cosh(2\pi\kappa l_k),\ l \in \mathcal{L}.$$

This corresponds to analytic functions f that provide periodic extensions to the complex d-dimensional strip $(t_1 + iu_1, \ldots, t_d + iu_d),\ |u_k| \le \kappa$ (i.e., of size 2κ), and

$$\|f\|_c^2 = 2^{-d} \sum_{\bar{\varepsilon}} \|f(\cdot + \varepsilon_k \kappa)\|^2.$$

This case is closely related to the case

$$c_l^2 = \exp\left(2\pi\kappa \sum_{k=1}^{d} |l_k|\right),\ l \in \mathcal{L}$$

(see [24]). Using $e^{|x|}/2 \le \cosh(x) \le e^{|x|}$, condition (2.2) is fulfilled for any $\kappa > 0$ (by Remark 5.1), since

$$\sum_{l \in \mathcal{L}} 2^{J(l)} c_l^{-2} \le 2^d \sum_{l \in \mathcal{L}} c_l^{-2} \left(1 + 2\sum_{k=1}^{\infty} \exp(2\pi\kappa k)\right)^d < \infty.$$

Thus, assumption **(A3)** is fulfilled.

It was shown in [21] that

$$N(C) \sim \frac{2^d \log^d(C)}{(\pi\kappa)^d \Gamma(d+1)}.$$

Using equation (3.5), this yields

$$C \asymp \frac{n^{1/2}}{(\log(n))^{d/4}}.$$

Hence, assumptions (**B1**), (**B2**) are fulfilled; moreover $N(C)$ is a slowly varying function, i.e., assumption (**B3**) is also fulfilled. Thus, we get the separation rates

$$r_n^* = \frac{(\log(n))^{d/4}}{n^{1/2}},$$

and the sharp asymptotics

$$u_n^2 \sim \frac{(\pi\kappa)^d \Gamma(d+1) n^2 r_n^4}{2\log^d(n)}.$$

Assumption (**C1**) is thus fulfilled. Hence, we arrive at (2.1).

The last example corresponds to an infinitely dimensional extension of the ANOVA decomposition, that was first suggested to lift the curse of dimensionality in high-dimensional numerical integration (see [23], [28], [32]).

6.5 Sloan-Woźniakowski norm

Let $\Delta = [0,1]^\infty,\ \mathcal{L} = \mathbb{Z}_*^\infty$. Taking $\sigma > 0,\ s > 0$, let

$$c_l = \prod_{j \in \mathbb{N}\,:\, l_j \neq 0} j^s |2\pi l_j|^\sigma, \quad l \in \mathcal{L},\ s > 0,\ \sigma > 0,\ c_{0,\dots,0,\dots} = 1.$$

This corresponds to an infinite tensor product of weighed Hilbert spaces. Under an infinite-dimensional ANOVA expansion,

$$f(t) = \sum_u f_u(t_u), \qquad \int_\Delta f_u(t_u) dt_k = 0 \quad \forall\ k \in u,$$

where the sum is taken over all finite subsets $u \subset \mathbb{N}$, we obtain

$$\|f\|_c^2 = \sum_u \gamma(u) \|f_u\|_{c,u}^2, \quad \gamma(u) = \prod_{k \in u} k^{2s},$$

and, for $\sigma \in \mathbb{N}$, the norm $\|f_u\|_{c,u}^2$ of a 1-periodic f_u is determined by (6.7) (see [20] and compare with [23], [28], [32]).

Contrary to the previous examples, we are not aware of any embedding theorems for spaces of the Sloan-Woźniakowski type, and hence we cannot verify Assumption (**A3**) under minimal smoothness conditions (like $\sigma^* \stackrel{\Delta}{=} \min(\sigma, s) > 1/4$). However, condition (2.2), which leads to the Assumption (**A3**), is fulfilled for $\sigma^* > 1/2$. Indeed, let $(x_{k,j}),\ k \in \mathbb{Z}, 1 \le j \le d$, be a matrix. Applying the formula

$$\sum_{\bar{l} \in \mathbb{Z}^d} \prod_{j=1}^d x_{l_j,j} = \prod_{j=1}^d \sum_{l \in \mathbb{Z}} x_{k,j}, \quad \bar{l} = \{l_1, \dots, l_d) \in \mathbb{Z}^d,$$

to the matrix entries

$$x_{k,j} = \begin{cases} 1, & k = 0, \\ 2j^{-2s} |2\pi k|^{-2\sigma}, & k \neq 0, \end{cases}$$

and letting $d \to \infty$, we get, for $\sigma > 1/2$ and $s > 1/2$,

$$\begin{aligned} \sum_{l \in \mathcal{L}} 2^{J(l)} c_l^{-2} &= \sum_{l \in \mathcal{L}} \prod_{j \in \mathbb{N}\,:\, l_j \neq 0} 2j^{-2s} |2\pi l_j|^{-2\sigma} \\ &= \prod_{j \in \mathbb{N}} \left(1 + 2j^{-2s} \sum_{k \in \breve{\mathbb{Z}}} |2\pi k|^{-2\sigma} \right) < \infty; \quad \breve{\mathbb{Z}} = \mathbb{Z} \setminus \{0\}. \end{aligned}$$

Thus, by Remark 5.1, assumption **(A3)** is fulfilled for $\sigma^* > 1/2$.

For simplicity, we consider below only the case $\sigma \neq s$. It was shown in [20] that if $0 < \sigma < s$ then

$$N(C) \sim A_1 C^{1/\sigma} \exp(A_2 (\log C)^{\sigma/(\sigma+s)}) (\log C)^{-A_2},$$

and that if $0 < s < \sigma$ then

$$N(C) \sim B_1 C^{1/s} \exp(B_2 (\log C)^{1/2}) (\log C)^{-B_3},$$

where A_i, $i = 1, 2$, and B_i, $i = 1, 2, 3$, are positive constants which only depend on σ, s. Recall that $\sigma^* \triangleq \min(s, \sigma)$. Then, we get the following log-asymptotics

$$\log(N(C)) \sim \frac{\log(C)}{\sigma^*},$$

which correspond to the Sobolev norms for $d = 1$ and $\sigma = \sigma^*$.

It also follows that assumption (**B2**) is fulfilled while assumption (**B1**) is fulfilled for $\sigma^* > 1/4$. The separation rates are of the following form. If $0 < \sigma < s$, then

$$r_n^* \asymp n^{-2\sigma/(4\sigma+1)} \exp\left(C_1(\log(n))^{\sigma/(s+\sigma)}\right)(\log(n))^{-C_2},$$

and if $0 < s < \sigma$, then

$$r_n^* \asymp n^{-2s/(4s+1)} \exp\left(D_1\sqrt{\log(n)}\right)(\log(n))^{-D_2}.$$

These yield the following log-asymptotics

$$\log(r_n^*) \sim -\frac{2\sigma^* \log(n)}{4\sigma^* + 1}.$$

The sharp asymptotics are of the following form. If $0 < \sigma < s$, then

$$u_n^2 \sim C_3 n^2 r_n^{4+1/\sigma} \exp\left(-C_4(\log r_n^{-1})^{\sigma/(s+\sigma)}\right)(\log r_n^{-1})^{C_5}.$$

If $0 < s < \sigma$, then

$$u_n^2 \sim D_3 n^2 r_n^{4+1/s} \exp\left(-D_4\sqrt{\log r_n^{-1}}\right)(\log r_n^{-1})^{3/4},$$

where C_i, $i = 1, \ldots, 5$, and D_i, $i = 1, \ldots, 4$, are positive constants which only depend on σ, s. Thus, assumption (**C1**) is fulfilled. Hence, we arrive at (2.1).

7 Some General Remarks

In this section, we discuss how the main results, established in Theorems 1 and 2 (and, hence, Corollaries 1 and 2) can be extended to more general settings, involving non-uniform design schemes and unknown variances. Some remarks about adaptivity issues are also presented. We also present other, than the Fourier basis and its tensor product version, examples of basis functions that satisfy assumption (**A2**), and reveal how assumption (**A2**) can be replaced by a weaker assumption at the cost of replacing assumption (**B1**) with a slightly stronger assumption.

7.1 General random design schemes

The main results, established in Theorems 1 and 2, are evidently extended to random design points $y = (y^1, \ldots, y^d) \in \mathbb{R}^d$, $d \geq 1$, with a *known* product probability density function, $p(y) = p_1(y^1) \times \ldots \times p_d(y^d)$, by applying the coordinates Smirnov

transform, i.e., $y \to F(y) = (F_1(y^1), \ldots, F_d(y^d)) \in \Delta = [0,1]^d$, where F_k is the cumulative distribution function corresponding to the probability density function p_k. Indeed, consider the goodness-of-fit testing problem for testing the null hypothesis $H_0 : f = 0$ against the alternative $H_1 : f \in \mathcal{F}_P : \ \|f\|_{2,P} \geq r_n$, where $\mathcal{F}_P$ consists of functions defined on $\mathbb{R}^d$ and which have the form $g(y) = f(F(y))$, $y \in \mathbb{R}^d$, with $g \in \mathcal{F}$ and $\|f\|_{2,P} = (\int_{\mathbb{R}^d} f^2(y)p(y)d(y))^{1/2}$; note that, in this case, $\|f\|_{2,P} = \|g\|$. The corresponding test statistics are now based on the kernels (3.3) and (4.9) with $t = (t^1, \ldots, t^d)$ replaced by $F(y) = (F_1(y^1), \ldots, F_d(y^d))$ (compare with [15]).

We conjecture that the main results, established in Theorems 1 and 2, can be also extended, subject to some additional constraints similar to [15], to *unknown* product probability density functions by replacing $F(y) = (F_1(y^1), \ldots, F_d(y^d))$ with $F_n(y) = (F_{n,1}(y^1), \ldots, F_{n,d}(y^d))$ in the appropriate test statistics, where $F_{n,k}$ is the empirical distribution function corresponding to F_k for the design points $y_1^k, \ldots, y_n^k$; this development is, however, outside the scope of this paper.

7.2 Unknown variance

The results obtained in Theorems 1 and 2 are evidently true when $\xi_i \overset{iid}{\sim} \mathcal{N}(0,1)$ is replaced by $\xi_i \overset{iid}{\sim} \mathcal{N}(0,\tau^2)$, where τ^2 is a *known* variance with $0 < \tau^2 < \infty$, by multiplying u_n by the factor τ^{-2} and multiplying r_n^* by the factor τ, for the lower bounds, and by multiplying the kernels (3.3) and (4.9) by the factor τ^{-2}, for the upper bounds.

For an *unknown* variance τ^2 with $0 < \beta_1 \leq \tau^2 \leq \beta_2 < \infty$, we replace the multiplicative factor τ^{-2} appeared in the kernels (3.3) and (4.9) by τ_n^{-2}, where $\tau_n^2 = \sum_{i=1}^n x_i^2$. It is easily seen that

$$E_{n,f}\tau_n^2 = \tau^2 + \|f\|^2, \quad \mathrm{Var}_{n,f}\tau_n^2 = \frac{1}{n}(\|f\|_4^4 - \|f\|^4 + 4\tau^2\|f\|^2 + 2\tau^4) = o(1),$$

the latter being true from assumption **(A3)**. These yield $\tau_n^2 \sim (\tau^2 + \|f\|^2)$, in $P_{n,f}$-probability, which makes possible to repeat all the arguments presented in Appendix 2 (observe that, in Appendix 2, $\|f\|^2 = o(1)$ for "least favorable" alternative functions $f \in \mathcal{F}$).

The above observations indicated that the main results established in Theorems 1 and 2 still remain true when the variance τ^2 is either known or, when unknown, is replaced by an appropriate estimator as the one considered above.

7.3 Adaptivity

Typically, the smoothness parameter (σ for Sobolev norms, κ for analytic function, $\min(\sigma, s)$ for Sloan-Woźniakowski norms) is *unknown*. This leads to the so-called problem of *adaptivity*: one has to construct a test procedure that provides the best minimax efficiency (separation rates or sharp asymptotics) for a wide range of values of the unknown smoothness parameter. This problem was first studied in [29], and further developed in Chapter 7 of [18], for the 1-variable Gaussian white noise model. The idea is to use the Bonferroni procedure, i.e., to combine a collection of tests for a suitable grid in a region of the unknown smoothness parameter. It was shown in [18] and [29] that this procedure provides an asymptotically minimax adaptive testing with a small loss (one gets an additional (but unavoidable) $\log\log(\varepsilon^{-1})$ factor in the separation

rates). We conjecture that these ideas of adaptivity could be also developed for the multivariate nonparametric regression models considered in this paper but the exact details should be carefully addressed; this development is, however, outside the scope of this paper.

7.4 Other examples of basis functions satisfying Assumption (A2)

(a) (*Haar basis*): Let $\phi_{jk}(t)$, $j = 0, 1, \ldots,$ $k = 1, \ldots, 2^j$, $t \in [0,1]$, be the standard Haar orthonormal system on $[0,1]$ (see, e.g., Chapter 7 of [31]), where j is the scale parameter and k is the shift parameter. Note that, in this case, $\sum_k \phi_{jk}^2(t) = 2^j$, for each resolution j. Consider now the tensor product version of the Haar basis on $\Delta = [0,1]^d$, $d \geq 1$, and consider coefficients $c_l = c_j$, $l = ((j_1, k_1), \ldots, (j_d, k_d))$, that only depend on the scale parameter $j = (j_1, \ldots, j_d)$ and not on the shift parameter $k = (k_1, \ldots, k_d)$. Hence, by working along the lines of Section 5, it follows that the tensor product Haar basis functions on Δ satisfy Assumption **(A2)**.

(b) (*Walsh basis*): Let $\phi_j(t)$, $j = 0, 1, \ldots,$ $t \in [0,1]$, be the Walsh basis functions system on [0,1]; the Walsh basis functions take actually sums and differences of the Haar basis functions to obtain a complete orthonormal system (see, e.g., Chapter 7 of [31]). Note that, in this case, $|\phi_j(x)| = 1$, for each j. Consider now the tensor product version of the Walsh basis functions on $\Delta = [0,1]^d$, $d \geq 1$. Hence, it follows immediately that the tensor product Walsh basis functions on Δ satisfy Assumption **(A2)**.

(c) (*Orthonomal basis on a compact connected Riemannian manifold without boundary*): Let S be a compact connected Riemannian manifold without boundary and consider the orthonormal system of eigenfunctions $\phi_{jk}(x)$, $x \in S$, associated with the Laplacian (Laplace-Beltrami operator) on S, for different eigenvalues λ_j, $\lambda_1 < \lambda_2 < \ldots$ with $\lambda_j \to \infty$ as $j \to \infty$ (see, e.g., [4]). For each $j = 1, 2, \ldots,$ they satisfy the relation $\sum_{k=1}^{k_j} (\phi_{j,k}^2(x) - \mu^{-1}(S)) = 0$, where $k_j < \infty$ is the (algebraic) multiplicity of the eigenvalue λ_j and μ is the invariant measure on S (see, e.g., formula (3.18), p. 127 of [6], or the last line of p. 1256 of [4]). The above relation is a natural and deep extension of the classical relation $\sin^2(x) + \cos^2(x) = 1$ for the 1-dimensional circle. Similar to (a), consider now coefficients $c_{(j,k)} = c_j$ or corresponding coefficients $c_l = c_j$ for the tensor product basis functions on S^d, $d \geq 1$. Hence, by working along the lines of Section 5, it follows that the tensor product basis functions on S^d satisfy Assumption **(A2)**. Therefore, our general framework could be a platform to derive analogous statements to the ones given in Theorems 1 and 2 for minimax goodness-of-fit testing in nonparametric regression problems on compact connected Riemannian manifolds without boundary, S, or their products, S^d, but the details in the derivation of these statements should be carefully addressed; this development is, however, outside the scope of this paper.

7.5 Replacing assumption (A2) by a weaker assumption

Assumption **(A2)** can be replaced by the weaker assumption

$$\textbf{(A2a)} \qquad \sup_{t \in \Delta} \sum_{l \in \mathcal{N}(C)} \phi_l^2(t) = O(N(C)) \quad \text{as} \quad C \to \infty,$$

(it covers the cosines orthonormal system, compactly supported (other than the Haar basis) orthonormal wavelet systems, as well as their tensor product versions) by replacing assumption (**B1**) with the slightly stronger assumption

$$(\mathbf{B1a}) \qquad N = o(n^{2/3}).$$

Indeed, the only difference in the proofs of Theorems 1 and 2 is in the relation (8.9). In particular, one can use the Cauchy-Schwarz inequality which yields an additional factor N, and this is compensated by assumption (**B1a**).

8 Appendix 1: proof of lower bounds

Let us start with some extra notation. Recall first that $X_n = \{x_1, \ldots, x_n\}$, $T_n = \{t_1, \ldots, t_n\}$, $Z_n = (X_n, T_n)$, and $z_i = (x_i, t_i)$, and that $P_{n,f}$ is the probability measure that corresponds to Z_n whereas $E_{n,f}$ is the expectation over this probability measure. Denote also by $\mathrm{Var}_{n,f}$ the corresponding variance. Let $P_{n,T}$ be the probability measure that corresponds to T_n and $P^T_{n,f}$ be the conditional probability measure with respect to T_n. We denote by $E_{n,T}$ and $E^T_{n,f}$ the expectations over these probability measures, whereas $\mathrm{Var}_{n,T}$, $\mathrm{Var}^T_{n,f}$ are the corresponding variances. (Clearly, $E_{n,f}(\cdot) = E_{n,T}E^T_{n,f}(\cdot)$.) Also, for the function $f = \sum_l \theta_l \phi_l$, we denote the measure $P_{n,f}$ by $P_{n,\theta}$, with analogous notation for the expectations, conditional expectations and variances. Let also $E^{T,\xi}_n$ and $\mathrm{Var}^{T,\xi}_n$ be the expectation and variance of the conditional probability measure with respect to $\Xi_n = \{\xi_1, \ldots, \xi_n\}$, where $\xi_i \overset{iid}{\sim} \mathcal{N}(0,1)$. Certainly, $P_{n,\xi} = P_{n,0}$.

8.1 Lower bounds for Theorem 2

8.1.1 Priors

We use the constructions similar to [7] and follow, but with necessary modifications, techniques from [14]–[18]. It suffices to consider the case

$$u_n^2 \asymp 1. \tag{8.1}$$

Take $\delta \in (0,1)$, let $a_{l,n} = v_{l,n}(b,B)$ be the extremal collection for the extremal problem (4.1), (4.2) with $b = 1-\delta, B = 1+\delta$, and let $A = A_n$ be the diagonal matrix with diagonal elements $a_l = a_{l,n},\ l \in \mathcal{N}$.

Under (8.1), using (**C1**), (4.7), we have

$$u_n^2(b,B) = \frac{1}{2}\sum_{l\in\mathcal{N}} a^4_{l,n} \asymp 1, \quad D_n = N \max_{j\in\mathcal{N}} a^4_{j,n} \sim z_0^4 N \asymp 1. \tag{8.2}$$

Let $v = \sqrt{n}\theta$ and let $\pi_n(dv)$ be the Gaussian prior $\mathcal{N}(0, A^2)$ on the parametric space consisting of $\{v_l\}_{l\in\mathcal{L}} = \sqrt{n}\{\theta_l\}_{l\in\mathcal{L}}$, i.e., v_l are independent in l and, for each l, $v_l \sim \mathcal{N}(0, a_l^2)$ for $c_l < C$ and $v_l = 0$ for $c_l \geq C$, in π_n-probability.

Note that, in the sequence space of the "generalized" Fourier coefficients $\theta = \{\theta_l\}_{l\in\mathcal{L}}$ with respect to the orthonormal system $\{\phi_l\}_{l\in\mathcal{L}}$, the null hypothesis (1.2) (recall that

$f_0 = 0$) corresponds to $H_0 : \theta = 0$ and, assuming $f \in \mathcal{F}$, the alternative hypothesis (1.3) corresponds to

$$H_1 : \quad \sum_{l \in \mathcal{L}} c_l^2 \theta_l^2 \leq 1, \quad \sum_{l \in \mathcal{L}} \theta_l^2 \geq r_n^2. \tag{8.3}$$

Let $V_n = V_n(1, 1)$ be the set determined by (4.2) with $B = b = 1$; this corresponds to the alternative set (8.3).

Lemma 8.1 *For any $\delta \in (0, 1)$, one has $\pi_n(V_n) = 1 + o(1)$.*

Proof of Lemma 8.1. It follows from evaluations of π_n-expectations and variances of the random variables $\mathcal{H}_1(v) = \sum_{l \in \mathcal{N}} v_l^2$ and $\mathcal{H}_2 = \sum_{l \in \mathcal{N}} c_l^2 v_l^2$, and by using the Chebyshev inequality (compare with similar evaluations in [14], [17], [18]). $\square$

Let $\beta(P_{n,0}, P_{\pi_n}, \alpha)$ be the minimal type II error probability for a given level $\alpha \in (0, 1)$ and $\gamma(P_{n,0}, P_{\pi_n})$ be the minimal total error probability for testing the simple null hypothesis $H_0 : P = P_{n,0}$ against the simple Bayesian alternative $H_0 : P = P_{\pi_n}$ for the mixture $P_{\pi_n}(A) = \int P_{n,n^{-1/2}v}(A)\, \pi_n(dv)$. By Lemma 8.1 and using Proposition 2.11 in [18], we have

$$\beta(\mathcal{F}, r_n, \alpha) \geq \beta(P_{n,0}, P_{\pi_n}, \alpha) + o(1), \quad \gamma(\mathcal{F}, r_n) \geq \gamma(P_{n,0}, P_{\pi_n}) + o(1).$$

Hence, it suffices to show that

$$\beta(P_{n,0}, P_{\pi_n}, \alpha) \geq \Phi(H^{(\alpha)} - u_n) + o(1), \quad \gamma(P_{n,0}, P_{\pi_n}) \geq 2\Phi(-u_n/2) + o(1). \tag{8.4}$$

In order to obtain (8.4), it suffices to verify that, in $P_{n,0}$-probability,

$$\log(dP_{\pi_n}/dP_{n,0}) = -u_n^2/2 + u_n \zeta_n + \eta_n, \quad \eta_n \to 0, \quad \zeta_n \to \zeta \sim \mathcal{N}(0, 1) \tag{8.5}$$

(see [18], Section 4.3.1, formula (4.72)).

8.1.2 Likelihood ratio and correlation matrix

For $f(t) = \sum_{l \in \mathcal{N}} \theta_l \phi_l(t)$, the likelihood ratio is of the form

$$\frac{dP_{n,\theta}}{dP_{n,0}} = \frac{dP_{n,\theta}^T}{dP_{n,0}^T} = \exp\bigg(-\frac{1}{2} v' R v + \langle w, v \rangle_s \bigg), \quad \theta = \{\theta_l\}_{l \in \mathcal{N}}, \quad v = \sqrt{n}\theta,$$

where $w = \{w_l\}_{l \in \mathcal{N}}$, $w_l = w_{l,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \phi_l(t_i)$, and R is the correlation matrix

$$R = R_n = \{r_{jl}\}_{j,l \in \mathcal{N}}, \quad r_{jl} = \frac{1}{n} \sum_{i=1}^n \phi_j(t_i) \phi_l(t_i);$$

here, and in Section 9.1.3, $\langle \cdot, \cdot \rangle_s$ denotes the inner product in the sequence space.

Let $\mathrm{Tr}(\cdot)$ be the trace of a square matrix.

Lemma 8.2 *(1) The matrix R is symmetric and positively semi-defined. Moreover, $E_{n,T}R = I_N$, where $I_N = \{\delta_{jl}\}_{j,l\in\mathcal{N}}$ is the unit $N \times N$ matrix.*

*(2) Under (2.2) and (**B1**), one has*

$$E_{n,T}\mathrm{Tr}(R^2) \sim N, \tag{8.6}$$
$$E_{n,T}\mathrm{Tr}((R-I_N)^2) = o(N), \tag{8.7}$$
$$E_{n,T}\mathrm{Tr}(R^4) \sim N. \tag{8.8}$$

Proof of Lemma 8.2. First, we prove statement (1). For any $\tilde{x} = \{\tilde{x}_j\}_{j\in\mathcal{N}},\ \tilde{x}_j \in \mathbb{R}$, one has

$$\sum_{j,l\in\mathcal{N}} \tilde{x}_j\tilde{x}_l r_{jl} = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j\in\mathcal{N}}\tilde{x}_j\phi_j(t_i)\right)^2 \geq 0.$$

Since $\{\phi_l\}_{l\in\mathcal{N}}$ is an orthonormal system,

$$E_{n,T}r_{jl} = \int_\Delta \phi_j(t)\phi_l(t)dt = \delta_{jl}.$$

Thus, statement (1) follows.

Now, we prove statement (2). Analogously, we have, using (**A2**), (**B1**),

$$\begin{aligned} E_{n,T}(r_{jl}-\delta_{jl})^2 &= \mathrm{Var}_{n,T}r_{jl} = \frac{1}{n}\left(\int_\Delta \phi_j^2(t)\phi_l^2(t)dt - \delta_{jl}^2\right) \\ &= \frac{1}{n}\int_\Delta \phi_j^2(t)\phi_l^2(t)dt - \frac{1}{n}\delta_{jl} \end{aligned}$$

and

$$\begin{aligned} E_{n,T}\mathrm{Tr}((R-I_N)^2) &= \sum_{j,l\in\mathcal{N}} E_{n,T}(r_{jl}-\delta_{jl})^2 \leq \frac{1}{n}\int_\Delta \sum_{j,l\in\mathcal{N}}\phi_j^2(t)\phi_l^2(t)dt \\ &= \frac{1}{n}\int_\Delta\left(\sum_{j\in\mathcal{N}}\phi_j^2(t)\right)^2 dt = \frac{N^2}{n} = o(N), \end{aligned}$$

which yields (8.7). We obtain (8.6) from (8.7) since $\mathrm{Tr}(R^2) = \mathrm{Tr}((R-I_N)^2) + \mathrm{Tr}(I_N)$.

Let us now evaluate $E_{n,T}\mathrm{Tr}(R^4)$. Let $R^2 = \{b_{jl}\}_{j,l\in\mathcal{N}}$,

$$b_{jl} = \sum_{s\in\mathcal{N}} r_{js}r_{sl} = \frac{1}{n^2}\sum_{s\in\mathcal{N}}\sum_{\alpha,\beta=1}^{n}\phi_j(t_\alpha)\phi_s(t_\alpha)\phi_s(t_\beta)\phi_l(t_\beta).$$

We have

$$\begin{aligned} &\mathrm{Tr}(R^4) = \sum_{j,l\in\mathcal{N}} b_{jl}^2 \\ &= \frac{1}{n^4}\sum_{l,j,s,r\in\mathcal{N}}\sum_{\alpha,\beta,\gamma,\delta=1}^{n}\phi_j(t_\alpha)\phi_s(t_\alpha)\phi_s(t_\beta)\phi_l(t_\beta)\phi_j(t_\gamma)\phi_r(t_\gamma)\phi_r(t_\delta)\phi_l(t_\delta). \end{aligned}$$

Observe that

$$\begin{aligned}\sum_{\alpha,\beta,\gamma,\delta=1}^{n} & E_{n,T}\left\{\phi_j(t_\alpha)\phi_s(t_\alpha)\phi_s(t_\beta)\phi_l(t_\beta)\phi_j(t_\gamma)\phi_r(t_\gamma)\phi_r(t_\delta)\phi_l(t_\delta)\right\} \\ := \quad & S_4 + S_3 + S_2 + S_1,\end{aligned}$$

where S_4–S_1 correspond to the sums (we omit indexes j, l, r, s in notation of S_1–S_4)

$$\begin{aligned} S_4 &= 24 \sum_{1\le\alpha<\beta<\gamma<\delta\le n}, \\ S_3 &= 6\left(\sum_{1\le\alpha=\beta<\gamma<\delta\le n} + \sum_{1\le\alpha<\beta=\gamma<\delta\le n} + \sum_{1\le\alpha<\beta<\gamma=\delta\le n}\right), \\ S_2 &= 2\left(\sum_{1\le\alpha=\beta=\gamma<\delta\le n} + \sum_{1\le\alpha<\beta=\gamma=\delta\le n} + \sum_{1\le\alpha=\beta<\gamma=\delta\le n}\right), \\ S_1 &= \sum_{1\le\alpha=\beta=\gamma=\delta\le n}. \end{aligned}$$

By independence of t_i, and since $\{\phi_l\}$ is an orthonormal system, we have

$$\begin{aligned} S_4 &= C_4(n)\delta_{js}\delta_{sl}\delta_{jr}\delta_{rl}, \\ S_3 &= C_3(n)\Big\{\delta_{jr}\delta_{lr}\int_\Delta \phi_j(t)\phi_s^2(t)\phi_l(t)dt + \delta_{js}\delta_{rl}\int_\Delta \phi_s(t)\phi_l(t)\phi_j(t)\phi_r(t)dt \\ &+ \delta_{js}\delta_{sl}\int_\Delta \phi_j(t)\phi_r^2(t)\phi_l(t)dt\Big\}, \\ S_2 &= C_2(n)\Big\{\delta_{rl}\int_\Delta \phi_j^2(t)\phi_s^2(t)\phi_l(t)\phi_r(t)dt + \delta_{sj}\int_\Delta \phi_l^2(t)\phi_r^2(t)\phi_j(t)\phi_s(t)dt \\ &+ \left(\int_\Delta \phi_j(t)\phi_s^2(t)\phi_l(t)dt\right)\left(\int_\Delta \phi_j(u)\phi_r^2(u)\phi_l(u)du\right)\Big\}, \\ S_1 &= n\int_\Delta \phi_j^2(t)\phi_s^2(t)\phi_r^2(t)\phi_l^2(t)dt, \end{aligned}$$

where $C_4(n) \sim n^4,\ C_3(n) \asymp n^3,\ C_2(n) \asymp n^2$. Therefore,

$$\begin{aligned}
\frac{1}{n^4}\sum_{l,j,s,r\in\mathcal{N}} S_4 &= \frac{C_4(n)}{n^4}\sum_{l,j,s,r\in\mathcal{N}} \delta_{js}\delta_{sl}\delta_{jr}\delta_{rl} = \frac{NC_4(n)}{n^4} \sim N,\\
\frac{1}{n^4}\sum_{l,j,s,r\in\mathcal{N}} S_3 &= \frac{3C_3(n)}{n^4}\sum_{j,s\in\mathcal{N}} \int_\Delta \phi_j^2(t)\phi_s^2(t)dt\\
&= \frac{3C_3(n)}{n^4}\int_\Delta \left(\sum_{j\in\mathcal{N}}\phi_j^2(t)\right)^2 dt = \frac{3N^2C_3(n)}{n^4} = O(N^2/n),\\
\frac{1}{n^4}\sum_{l,j,s,r\in\mathcal{N}} S_1 &= \frac{n}{n^4}\sum_{l,j,s,r\in\mathcal{N}} \int_\Delta \phi_j^2(t)\phi_s^2(t)\phi_r^2(t)\phi_l^2(t)dt\\
&= \frac{1}{n^3}\int_\Delta \left(\sum_{j\in\mathcal{N}}\phi_j^2(t)\right)^4 dt = \frac{N^4}{n^3}.
\end{aligned}$$

Analogously,

$$\begin{aligned}
&\sum_{l,j,s,r\in\mathcal{N}} \delta_{rl}\int_\Delta \phi_j^2(t)\phi_s^2(t)\phi_l(t)\phi_r(t)dt = \int_\Delta \left(\sum_{l,j,s\in\mathcal{N}} \phi_j^2(t)\phi_s^2(t)\phi_l^2(t)\right) dt\\
&= \int_\Delta \left(\sum_{l\in\mathcal{N}}\phi_j^2(t)\right)^3 dt = N^3
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{l,j,s,r\in\mathcal{N}} \left(\int_\Delta \phi_j(t)\phi_s^2(t)\phi_l(t)dt\right)\left(\int_\Delta \phi_j(u)\phi_r^2(u)\phi_l(u)du\right)\\
&= \sum_{l,j\in\mathcal{N}} \left(\int_\Delta \phi_j(t)\left(\sum_{s\in\mathcal{N}}\phi_s^2(t)\right)\phi_l(t)dt\right)\left(\int_\Delta \phi_j(u)\left(\sum_{s\in\mathcal{N}}\phi_r^2(u)\right)\phi_l(u)du\right) \qquad (8.9)\\
&= N^2\sum_{l,j\in\mathcal{N}} \left(\int_\Delta \phi_j(t)\phi_l(t)dt\right)\left(\int_\Delta \phi_j(u)\phi_l(u)du\right) = N^2\sum_{l,j\in\mathcal{N}}\delta_{jl}^2 = N^3.
\end{aligned}$$

Thus,

$$\frac{1}{n^4}\sum_{l,j,s,r\in\mathcal{N}} S_2 = O(N^3/n^2).$$

Combining evaluations above and (**B1**) we get (8.8):

$$\mathrm{Tr}(R^4) \sim N(1 + O(N/n + (N/n)^2 + (N/n)^3)) \sim N.$$

Thus, statement (2) follows. This competes the proof of Lemma 8.2. $\square$

8.1.3 Bayesian likelihood ratio

Let us now study the Bayesian likelihood ratio. Direct calculation gives

$$\frac{dP_{\pi_n}}{dP_{n,0}} = E_{\pi_n}\frac{dP^T_{n,\theta}}{dP^T_{n,0}} = \frac{1}{\sqrt{\det G}}\exp\left(\frac{1}{2}q'G^{-1}q\right), \tag{8.10}$$

where $q = Aw,\ G = G_n = I_N + A'RA$. Let $\tilde{\tau}_l \geq 0,\ l \in \mathcal{N}$, be the eigenvalues of the symmetric positively semi-defined matrix $D = A'RA = \{a_j a_l r_{jl}\}_{j,l\in\mathcal{N}}$. Let e_l be the eigenvectors of the matrix D and let $q_l = \langle q, e_l\rangle_s$, $l \in \mathcal{L}$.

We can now rewrite (8.10) in the form

$$L_n = \log\left(\frac{dP_{\pi_n}}{dP_{n,0}}\right) = \frac{1}{2}\sum_{l\in\mathcal{N}}\left(\frac{q_l^2}{1+\tilde{\tau}_l} - \log(1+\tilde{\tau}_l)\right).$$

Let $\|\tilde{A}\|_\infty = \sup_{\|x\|\leq 1}\|\tilde{A}x\|$ for a generic matrix $\tilde{A}$. Observe that

$$\|D\|_\infty^4 = \max_{l\in\mathcal{N}}\tilde{\tau}_l^4 \leq \sum_{l\in\mathcal{N}}\tilde{\tau}_l^4 = \mathrm{Tr}(D^4).$$

Using the standard relations

$$\mathrm{Tr}(AC) = \mathrm{Tr}(CA) \quad \text{and} \quad \mathrm{Tr}(A'BA) \leq \|A\|_\infty^2\,\mathrm{Tr}(B),$$

for a symmetric positively semi-defined matrix B, we get the inequalities

$$\mathrm{Tr}(D^2) \leq \|A\|_\infty^4\,\mathrm{Tr}(R^2) \quad \text{and} \quad \mathrm{Tr}(D^4) \leq \|A\|_\infty^8\,\mathrm{Tr}(R^4).$$

By (8.2),

$$\|A\|_\infty^4 = \max_{l\in\mathcal{N}} a_l^4 \leq D_n/N.$$

Jointly with (8.6) and (8.8), the above yields

$$E_{n,T}(\mathrm{Tr}(D^2)) = O(1), \quad E_{n,T}(\mathrm{Tr}(D^4)) = O(N^{-1}).$$

Hence,

$$E_{n,T}\left(\max_{l\in\mathcal{N}}|\tilde{\tau}_l|\right) = O(N^{-1/4}).$$

Thus, in $P_{n,T}$-probability,

$$\|D\|_\infty = \max_{l\in\mathcal{N}}|\tilde{\tau}_l| = o(1). \tag{8.11}$$

Using the well-known relations

$$(1+y)^{-1} = 1 - y + o(y) \quad \text{and} \quad \log(1+y) - y + y^2/2 = o(y^2), \quad \text{as} \quad y \to 0,$$

we get, with $P_{n,T}$-probability tending to 1, by (8.11),

$$\begin{aligned} L_n &= \frac{1}{2}\sum_{l\in\mathcal{N}}\left(q_l^2(1-\tilde{\tau}_l) - \tilde{\tau}_l + \tilde{\tau}_l^2/2\right) + o\left(\sum_{l\in\mathcal{N}} q_l^2\tilde{\tau}_l\right) + o\left(\sum_{l\in\mathcal{N}}\tilde{\tau}_l^2\right) \\ &= \frac{1}{2}\left(\mathrm{Tr}(Q) - \mathrm{Tr}(D) - \mathrm{Tr}(QD) + \mathrm{Tr}(D^2)/2\right) + o\left(\mathrm{Tr}(QD)\right) + o\left(\mathrm{Tr}(D^2)\right) \\ &= \frac{1}{2}\left(\mathrm{Tr}(\hat{Q}) - \mathrm{Tr}(\hat{Q}D) - \mathrm{Tr}(D^2)/2\right) + o\left(\mathrm{Tr}(\hat{Q}D)\right) + o\left(\mathrm{Tr}(D^2)\right), \end{aligned} \tag{8.12}$$

where

$$Q = qq' = Azz'A = \{a_j a_l z_j z_l\}_{j,l\in\mathcal{N}},\ \hat{Q} = Q - D = A(zz' - R)A.$$

Let us now study the $P_{n,0}$-distribution of L_n.

Lemma 8.3 *In $P_{n,0}$-probability,*

$$\begin{aligned} \mathrm{Tr}(\hat{Q}D) &= o(1), & (8.13)\\ \mathrm{Tr}(D^2) &= \mathrm{Tr}(A^4) + o(1), & (8.14)\\ E_{n,0}\mathrm{Tr}(\hat{Q}) &= 0, & (8.15)\\ \mathrm{Var}_{n,0}\mathrm{Tr}(\hat{Q}) &= 2\mathrm{Tr}(A^4) + o(1). & (8.16) \end{aligned}$$

Proof of Lemma 8.3. Let $\Phi = n^{-1/2}\{\phi_j(t_i)\}_{j\in\mathcal{N},i=1,\ldots,n}$ be an $N \times n$-matrix, and set $\xi' = (\xi_1, \ldots, \xi_n)$. Then, in $P_{n,0}$-probability,

$$R = \Phi\Phi', \quad z = \Phi\xi, \quad z'z = \xi'\Phi'\Phi\xi,\ E(\xi\xi') = I_N.$$

Observe that

$$E^T_{n,0}zz' = \Phi\left(E^T_{n,0}\xi\xi'\right)\Phi' = \Phi\Phi' = R,$$

which yields

$$E^T_{n,0}(\mathrm{Tr}(\hat{Q})) = 0, \quad E^T_{n,0}(\mathrm{Tr}(\hat{Q}D)) = 0. \tag{8.17}$$

Analogously, using the formula

$$\mathrm{Var}(\mathrm{Tr}(B\xi\xi'))) = 2\mathrm{Tr}(BB'),$$

we get

$$\mathrm{Var}^T_{n,0}(\mathrm{Tr}(\hat{Q}D)) = \mathrm{Var}^T_{n,0}\mathrm{Tr}(A\Phi\xi\xi'\Phi'AD) = 2\mathrm{Tr}(BB'),$$

where $B = \Phi'A^2\Phi\Phi'A^2\Phi$. By Lemma 8.2 and (8.2), it is easily seen that

$$\mathrm{Tr}(BB') = \mathrm{Tr}((ARA)^4) \le \|A\|^8_\infty \mathrm{Tr}(R^4).$$

Using the formula

$$\mathrm{Var}_{n,0}(\cdot) = \mathrm{Var}_T(E^T_{n,0}(\cdot)) + E_T(\mathrm{Var}^T_{n,0}(\cdot)),$$

we get

$$\mathrm{Var}_{n,0}(\mathrm{Tr}(\hat{Q}D)) = o(1),$$

which together with (8.17), yields (8.13).

To obtain (8.14), note that

$$\mathrm{Tr}(D^2) = \mathrm{Tr}(\hat{D}^2) + 2\mathrm{Tr}(A^2\hat{D}) + \mathrm{Tr}(A^4), \quad \hat{D} = D - A^2 = A(R - I_N)A,$$

and observe that, by Lemma 8.2 and (8.2),

$$\mathrm{Tr}(\hat{D}^2) \le \|A\|^4_\infty \mathrm{Tr}((R - I_N)^2) = o(1), \quad (\mathrm{Tr}(A^2\hat{D}))^2 \le \mathrm{Tr}(A^4)\mathrm{Tr}(\hat{D}^2) = o(1).$$

Obviously, (8.15) follows from (8.17), and (8.16) follows from (8.14), since

$$\mathrm{Var}^T_{n,0}(\mathrm{Tr}(\hat{Q})) = \mathrm{Var}^T_{n,0}(\mathrm{Tr}(A\Phi\xi\xi'\Phi'A)) = 2\mathrm{Tr}((A\Phi\Phi'A)^2) = 2\mathrm{Tr}(D^2).$$

This completes the proof of Lemma 8.3. □

Let $\zeta_n = \mathrm{Tr}(\hat{Q})/2u_n,\ u^2_n = \mathrm{Tr}(A^4)/2$. By Lemma 8.3, we rewrite (8.12) in the form

$$L_n = u_n\zeta_n - u^2_n/2 + \eta_n, \quad \eta_n \overset{P_{n,0}}{\to} 0.$$

Lemma 8.4 *In $P_{n,0}$-probability, $\zeta_n \to \zeta \sim \mathcal{N}(0,1)$.*

Proof of Lemma 8.4. Let us rewrite $\mathrm{Tr}(\hat{Q})$ in the form

$$\begin{aligned}\frac{1}{2}\mathrm{Tr}(\hat{Q}) &= \frac{1}{2}\mathrm{Tr}(A\Phi(\xi\xi'-I)\Phi'A) = \frac{1}{2}\sum_{i=1}^{n} w_{ii}(\xi_i^2-1) + \sum_{1\le i<k\le n} w_{ik}\xi_i\xi_j \\ &:= A_n + B_n,\end{aligned}$$

where

$$W = \{w_{ik}\}_{i,k=1}^{n} = \Phi' A^2 \Phi, \quad w_{ik} = \frac{1}{n}\sum_{l\in\mathcal{N}} a_l^2 \phi_l(t_i)\phi_l(t_k).$$

It is easily seen that $E_n^{T,\xi} A_n = 0$, and by (**A2**), (8.2),

$$\begin{aligned}\mathrm{Var}_n^{T,\xi}(A_n) &= \frac{1}{2}\sum_{i=1}^{n} w_{ii}^2 = \frac{1}{2n^2}\sum_{i=1}^{n}\left(\sum_{l\in\mathcal{N}} a_l^2\phi_l^2(t_i)\right)^2 \\ &\le \frac{D_n}{2n^2 N}\sum_{i=1}^{n}\left(\sum_{l\in\mathcal{N}}\phi_l^2(t_i)\right)^2 = \frac{D_n N}{2n} = o(1).\end{aligned}$$

Thus, $A_n \to 0$ in $L_2(P_{n,0})$ and in $P_{n,0}$-probability.

The item B_n is degenerate U-statistic

$$B_n = \frac{1}{n}\sum_{1\le i<k\le n} W_n(r_i, r_j), \quad r_i = (\xi_i, t_i) \quad \text{are} \quad i.i.d.,$$

$$W_n(r', r'') = \xi'\xi'' \sum_{l\in\mathcal{N}} a_l^2 \phi_l(t')\phi_l(t''), \qquad \int W_n(r', r'') P(dr') = 0 \quad \forall r'',$$

where $P(dr) = \mathcal{N}_{0,1}(d\xi) \times U_\Delta(dt)$, i.e., ξ and t are independent, $\xi \sim \mathcal{N}(0,1)$ and t is uniformly distributed on Δ.

The statement of Lemma 8.4 follows from the following proposition.

Proposition 1 *In $P_{n,0}$-probability, the statistics B_n are asymptotically $\mathcal{N}(0, u_n^2)$.*

Proof of Proposition 1. Clearly, $E_{P_{n,0}} B_n = 0$ and, for $r_1 = (\xi_1, t_1),\ r_2 = (\xi_2, t_2)$,

$$\begin{aligned}\mathrm{Var}_{P_{n,0}}(B_n) &= \frac{n(n-1)}{2n^2}\int\int W_n^2(r_1, r_2) P(dr_1) P(dr_2) \\ &= \frac{n(n-1)}{2n^2} E(\xi_1^2\xi_2^2)\int_\Delta\int_\Delta\left(\sum_{l\in\mathcal{N}} a_l^2\phi_l(t_1)\phi_l(t_2)\right)^2 dt_1 dt_2 \\ &= \frac{n(n-1)}{2n^2}\sum_{j,l\in\mathcal{N}} a_j^2 a_l^2 \int_\Delta\int_\Delta \phi_j(t_1)\phi_j(t_2)\phi_l(t_1)\phi_l(t_2) dt_1 dt_2 \\ &= \frac{n(n-1)}{2n^2}\sum_{l\in\mathcal{N}} a_l^4 \sim u_n^2.\end{aligned}$$

For $r_1 = (\xi_1, t_1),\ r_2 = (\xi_2, t_2),\ r_3 = (\xi_3, t_3)$, let

$$
\begin{aligned}
\widetilde{G}_n(r_1, r_2) &= \int W_n(r_1, r_3) W_n(r_2, r_3) P(dr_3), \\
G_{n,2} &= \int\int \widetilde{G}_n^2(r_1, r_2) P(dr_1) P(dr_2), \\
W_{n,4} &= \int\int W_n^4(r_1, r_2) P(dr_1) P(dr_2).
\end{aligned}
$$

Using the asymptotic normality of degenerate U-statistics established in [10], together with Lemma 3.4 in [16], it suffices to verify the conditions

$$
\begin{aligned}
\widetilde{G}_{n,2} &= o(1), && (8.18)\\
W_{n,4} &= o(n^2). && (8.19)
\end{aligned}
$$

We have

$$
\begin{aligned}
\widetilde{G}_n(r_1, r_2) &= E_{P(d\xi_3, dt_3)}\left(\xi_1\xi_2\xi_3^2 \sum_{l\in\mathcal{N}} a_l^2 \phi_l(t_1)\phi_l(t_3) \sum_{j\in\mathcal{N}} a_j^2 \phi_j(t_2)\phi_j(t_3)\right) \\
&= \xi_1\xi_2 \sum_{j,l\in\mathcal{N}} a_l^2 a_j^2 \phi_l(t_1)\phi_j(t_2) \int_\Delta \phi_l(t_3)\phi_j(t_3) dt_3 = \xi_1\xi_2 \sum_{l\in\mathcal{N}} a_l^4 \phi_l(t_1)\phi_l(t_2), \\
G_{n,2} &= E(\xi_1\xi_2)^2 \int_\Delta\int_\Delta \left(\sum_{l\in\mathcal{N}} a_l^4 \phi_l(t_1)\phi_l(t_2)\right)^2 dt_1 dt_2 = \sum_{l\in\mathcal{N}} a_l^8 = O(N^{-1}),
\end{aligned}
$$

which yields (8.18). Next,

$$
\begin{aligned}
W_{n,4} &= E(\xi_1\xi_2)^4 \int_\Delta\int_\Delta \left(\sum_{l\in\mathcal{N}} a_l^2 \phi_l(t_1)\phi_l(t_2)\right)^4 dt_1 dt_2 \\
&\le 9 \sup_{t_1,t_2\in\Delta} \left(\sum_{l\in\mathcal{N}} a_l^2 \phi_l(t_1)\phi_l(t_2)\right)^2 \int_\Delta\int_\Delta \left(\sum_{l\in\mathcal{N}} a_l^2 \phi_l(t_1)\phi_l(t_2)\right)^2 dt_1 dt_2 = O(N),
\end{aligned}
$$

since by (**A2**) and (8.2), we have

$$
\sup_{t_1,t_2\in\Delta} \left|\sum_{l\in\mathcal{N}} a_l^2 \phi_l(t_1)\phi_l(t_2)\right| = \sup_{t_1\in\Delta} \sum_{l\in\mathcal{N}} a_l^2 \phi_l^2(t_1) \le \max_{l\in\mathcal{N}} a_l^2 \sup_{t_1\in\Delta} \sum_{l\in\mathcal{N}} \phi_l^2(t_1) = O(N^{1/2}).
$$

This implies (8.19). This completes the proof of Proposition 1. Hence, Lemma 8.4 follows. □

Thus, we obtain (8.5) which yields (8.4). Hence, Theorem 2 (1) follows. □

8.2 Lower bounds for Theorem 1

The same scheme used in the proof of the lower bounds of Theorem 2 can be also employed here.

Let $C^2 r_n^2 < (1-\delta)$, $\delta > 0$. It suffices to assume $u_n^2 = n^2 r_n^4/2N = O(1)$. We take the Gaussian prior $\pi_n = \mathcal{N}(0, A^2)$ that corresponds to the matrix $A = a_n I_N$ with $a_n^2 = n r_n^2 (1+\delta)/N$. Recall $\mathcal{H}_1$, $\mathcal{H}_2$ from the proof of Lemma 8.1. Analogously to the proof of Lemma 8.1, we have

$$\begin{aligned}
E_{\pi_n}\mathcal{H}_1 &= a_n^2 N = n r_n^2 (1+\delta),\\
E_{\pi_n}\mathcal{H}_2 &\leq C^2 a_n^2 N < n C^2 r_n^2 (1-\delta) < n,\\
\mathrm{Var}_{\pi_n}\mathcal{H}_1 &= 2 a_n^4 N = O(1),\\
\mathrm{Var}_{\pi_n}\mathcal{H}_2 &\leq 2 C^4 a_n^4 N = O(n^2/N).
\end{aligned}$$

Since, by Chebyshev's inequality, $\mathrm{Var}_{\pi_n}\mathcal{H}_k = o((E_{\pi_n}\mathcal{H}_k)^2)$, $k = 1, 2$, these yields $\pi_n(V_n) = 1 + o(1)$.

Observe that the relations (8.2) hold true with $z_0 = a_n$. Repeating the calculations in the proof of the lower bounds of Theorem 2, we arrive at (8.4) with $u_n^2 = N a_n^4/2 = n^2 r_n^4 / 2N(1+\delta)^2$. Since $\delta > 0$ can be taken arbitrary small, this yields Theorem 1 (1). □

9 Appendix 2: proof of upper bounds

9.1 Upper bounds for Theorem 2

We consider the test sequence $\psi_n^H = \mathbb{1}_{\{U_n > H\}}$ based on the U-statistics U_n with the kernel $K_n(z_1, z_2)$ of the form (4.9).

9.1.1 Type I error

Observe that $K_n(z_1, z_2) = u_n^{-1} W_n(z_1, z_2)$, where W_n is the kernel of the U-statistics mentioned in Proposition 1. Applying Proposition 1, we get

$$U_n \overset{P_{n,0}}{\to} \zeta \sim \mathcal{N}(0,1).$$

This yields,

$$E_{n,0}(\psi_n^H) = P_{n,0}(U_n \leq -H) = 1 - \Phi(H) + o(1). \tag{9.1}$$

9.1.2 Minimax type II error

By (9.1) we have to verify that

$$\sup_{f \in \mathcal{F}(r_n)} E_{n,f}(1 - \psi_n^H) = \sup_{f \in \mathcal{F}(r_n)} P_{n,f}(U_n > H) = \Phi(H - u_n) + o(1). \tag{9.2}$$

For $f = \sum_{l \in \mathcal{L}} \theta_l \phi_l$, let

$$v_l = \sqrt{n}\theta_l, \quad h_n(f) = \frac{1}{2}\sum_{l \in \mathcal{N}} w_{n,l} v_l^2.$$

Lemma 9.1 *Uniformly over $f \in \mathcal{F}$,*

$$E_{n,f}U_n \sim h_n(f), \tag{9.3}$$

$$\mathrm{Var}_{n,f}U_n = 1 + O(\|f\|^2 + \|f\|_4^4). \tag{9.4}$$

Moreover, uniformly over $f \in \mathcal{F}$ such that

$$\|f\| = o(1), \quad \|f\|_4 = o(1) \quad \textit{and} \quad h_n(f) = O(1), \tag{9.5}$$

the statistics $U_n - h_n(f)$ are asymptotically $\mathcal{N}(0,1)$, under $P_{n,f}$-probability.

Remark 9.1 Using Hölder's inequality and (**A3**) with $p = 4 + 2\delta,\ \delta > 0$, we get

$$\|f\|_4^4 \le \|f\|^a \|f\|_p^b, \quad a = 2/(1+1/\delta), \quad b = p/(1+\delta); \quad \|f\| \le \|f\|_p.$$

Therefore, under (**A3**), Lemma 9.1 yields

$$\sup_{f\in\mathcal{F}} \mathrm{Var}_{n,f}U_n = O(1) \quad \text{and} \quad \mathrm{Var}_{n,f}U_n = 1 + O(\|f\|^2 + \|f\|^a) \tag{9.6}$$

uniformly over $f \in \mathcal{F}$, and

$$U_n = h_n(f) + \zeta_n,\ \zeta_n \to \zeta \sim \mathcal{N}(0,1),$$

uniformly over $f \in \mathcal{F}$ such that $h_n(f) = O(1)$ and $\|f\| = o(1)$.

Proof of Lemma 9.1. Let the function $f = n^{-1/2}\sum_{l\in\mathcal{L}} v_l\phi_l$. Denote $z = (x,t)$ with $x = f(t) + \xi$, ξ and t are independent, $\xi \sim \mathcal{N}(0,1)$ and t is uniformly distributed on Δ. Since the items of the sum in U-statistics are identically distributed and uncorrelated, we have

$$E_{n,f}U_n = \frac{n-1}{2}E_{n,f}K_n(z_1,z_2),$$

where z_1 and z_2 are independent and distributed as z,

$$\begin{aligned} E_{n,f}K_n(z_1,z_2) &= E_{n,f}x_1x_2G_n(t_1,t_2) = E_n^T f(t_1)f(t_2)G_n(t_1,t_2) \\ &= \sum_{l\in\mathcal{N}} w_{n,l}E_n^T\,(f(t)\phi_l(t))^2 = n^{-1}\sum_{l\in\mathcal{N}} w_{n,l}v_l^2. \end{aligned}$$

Hence, (9.3) follows.

Let us now evaluate the variance. Rewrite the U-statistics in the form

$$U_n = U_{n,0} + U_{n,1} + U_{n,2}, \tag{9.7}$$

where

$$U_{n,k} = \frac{1}{n}\sum_{1\le i<j\le n} K_{n,k}(z_i,z_j)$$

are U-statistics with the kernels $K_{n,k}(z_1,z_2)$ of the form

$$\begin{aligned} K_{n,0} &= \xi_1\xi_2G_n(t_1,t_2), \quad K_{n,1} = (\xi_1 f(t_2) + \xi_2 f(t_1))G_n(t_1,t_2), \\ K_{n,2} &= f(t_1)f(t_2)G_n(t_1,t_2), \quad G_n(t_1,t_2) = \sum_{l\in\mathcal{N}} w_{n,l}\phi_l(t_1)\phi_l(t_2), \end{aligned}$$

and the items $U_{n,0}$, $U_{n,1}$ and $U_{n,2}$ are uncorrelated. Obviously,

$$\begin{aligned}E_{n,f}U_{n,0} &= E_{n,f}U_{n,1} = 0,\\ E_{n,f}U_{n,2} &= \frac{n-1}{2}\sum_{l\in\mathcal{N}} w_{n,l}\left(\int_\Delta f(t)\phi_l(t)dt\right)^2 \sim h_n(f).\end{aligned}$$

Similarly to Proposition 1,

$$\mathrm{Var}_{n,f}U_{n,0} \sim \frac{1}{2}\int_\Delta\int_\Delta G_n^2(t_1,t_2)dt_1dt_2 = \frac{1}{2}\sum_{l\in\mathcal{N}} w_{n,l}^2 = 1.$$

Analogously, by (**A2**) and (4.7), and since $\max_l w_{n,l}^2 = O(1/N)$,

$$\begin{aligned}\mathrm{Var}_{n,f}U_{n,1} &\sim 2\int_\Delta\int_\Delta f^2(t_1)G_n^2(t_1,t_2)dt_1dt_2\\ &= 2\int_\Delta\left(f^2(t)\sum_{l\in\mathcal{N}} w_{n,l}^2\phi_l^2(t)\right)dt = O(\|f\|^2).\end{aligned}$$

Next,

$$\mathrm{Var}_{n,f}U_{n,2} \le \int_\Delta\int_\Delta f^2(t_1)f^2(t_2)G_n^2(t_1,t_2)dt_1dt_2 = A_n.$$

Let $\mathbf{G}_n$ be the integral operator in $L_2(\Delta)$ associated with the symmetric positively semi-defined kernel $G_n(t_1,t_2),\ t_1,t_2\in\Delta$, and

$$\|\mathbf{G}_n\|_\infty = \sup_{\|f\|\le 1}\|\mathbf{G}_n f\| = \max_{l\in\mathcal{N}} w_{n,l} = O(N^{-1/2}).$$

Observe that, by (**A2**) and (4.7),

$$\mathbf{G}_n^* = \sup_{t\in\Delta}\sum_{l\in\mathcal{N}} w_{n,l}\phi_l^2(t) \le N\|\mathbf{G}_n\|_\infty, \quad \mathbf{G}_n^*\|\mathbf{G}_n\|_\infty = O(1).$$

We have

$$\begin{aligned}A_n &= \sum_{l\in\mathcal{N}} w_{n,l}\int_\Delta\int_\Delta \phi_l(t_1)\phi_l(t_2)f^2(t_1)f^2(t_2)G_n(t_1,t_2)dt_1dt_2\\ &= \sum_{l\in\mathcal{N}} w_{n,l}\langle f^2\phi_l, \mathbf{G}_n(f^2\phi_l)\rangle \le \|\mathbf{G}_n\|_\infty\sum_{l\in\mathcal{N}} w_{n,l}\|f^2\phi_l\|^2\\ &= \|\mathbf{G}_n\|_\infty\int_\Delta\sum_{l\in\mathcal{N}} w_{n,l}\phi_l^2(t)f^4(t)dt\\ &\le \|\mathbf{G}_n\|_\infty\sup_{t\in\Delta}\left(\sum_{l\in\mathcal{N}} w_{n,l}\phi_l^2(t)\right)\int_\Delta f^4(t)dt\\ &= \|\mathbf{G}_n\|_\infty\mathbf{G}_n^*\|f\|_4^4 = O(\|f\|_4^4).\end{aligned}$$

Hence, (9.4) follows.

Using (9.7), and an evaluation similar to the above under (9.5), we have

$$U_n - h_n(f) = U_{n,0} + U_{n,1} + U_{n,2} - h_n(f),$$

where $U_{n,1} \to 0$, $U_{n,2} - h_n(f) \to 0$, in $P_{n,f}$-probability. By Proposition 1, the statistics $U_{n,0}$ are asymptotically Gaussian $\mathcal{N}(0,1)$. This completes the proof of Lemma 9.1. □

Let $h_n(f) = O(1)$. Let us now evaluate $\|f\|^2$, $f \in \mathcal{F}$. We have

$$\|f\|^2 = \sum_{l\in\mathcal{L}} \theta_l^2 := A_n' + B_n', \quad A_n' = \sum_{c_l < C/2} \theta_l^2, \quad B_n' = \sum_{c_l \geq C/2} \theta_l^2.$$

The second sum is controlled by

$$B_n' \leq 4C^{-2} \sum_{l\in\mathcal{L}} c_l^2 \theta_l^2 \leq 4C^{-2} = o(1).$$

The first sum is controlled by

$$\begin{aligned} A_n' &\leq (4/3) \sum_{l\in\mathcal{N}} (1-(c_l/C)^2)\theta_l^2 = (4/3)(w_n/n) \sum_{l\in\mathcal{N}} w_{n,l} v_n^2 \\ &= (4/3)(w_n/n) h_n(f) = o(h_n(f)), \end{aligned}$$

since, by (4.7) and (**B1**), we have $w_n/n = O(N^{1/2}/n) = o(1)$. Therefore, by (9.6), we have in $P_{n,f}$-probability,

$$U_n = h_n(f) + \zeta_n, \quad \zeta_n \to \zeta \sim \mathcal{N}(0,1),$$

uniformly as $h_n(f) = O(1)$.

Lemma 9.2

$$\inf_{f\in\mathcal{F}(r_n)} h_n(f) = u_n.$$

Proof of Lemma 9.2 It follows using general convexity arguments (see [14], Lemma 11 of [17], Proposition 4.1 of [18]). □

Let us now evaluate type II errors for a sequence $f = f_n \in \mathcal{F}(r_n)$. First, let $h_n(f_n) \to \infty$. Applying Lemmas 9.1, 9.2 and (9.6), we have

$$\begin{aligned} E_{n,f}(1-\psi_n^H) &= P_{n,f}(U_n \leq H) = P_{n,f}(E_{n,f} - U_n \geq E_{n,f} - H) \\ &\leq \mathrm{Var}_{n,f}(U_n)/(E_{n,f} - H)^2 = o(1). \end{aligned}$$

Let $h_n(f_n) = O(1)$ (by Lemma 9.2 this is only possible for $u_n = O(1)$). Applying Lemmas 9.1, 9.2 and (9.6) once again, we have

$$\begin{aligned} E_{n,f}(1-\psi_n^H) &= P_{n,f}(U_n \leq H) = P_{n,f}(E_{n,f} - U_n \geq E_{n,f} - H) \\ &= P_{n,f}(\zeta_n \geq h_n(f) - H + o(1)) = \Phi(H - h_n(f)) + o(1). \end{aligned}$$

Therefore,

$$\sup_{f\in\mathcal{F}(r_n)} E_{n,f}(1-\psi_n^H) = \Phi(H - \inf_{f\in\mathcal{F}(r_n)} h_n(f)) + o(1) = \Phi(H - u_n) + o(1).$$

This yields (9.2). Hence, Theorem 2 (2) follows. □

This completes the proof of Theorem 2.

9.2 Upper bounds for Theorem 1

Observe that the kernel (3.3) is of the form (4.9) with coefficients

$$w_{l,n} = w_n = \sqrt{2/N}, \quad l \in \mathcal{N}.$$

Hence, Proposition 1 is applicable to the U-statistics U_n with kernel (3.3) and yields asymptotic normality $\mathcal{N}(0,1)$ of U_n under $P_{n,0}$. Thus, we get (9.1). Analogously, we obtain Lemma 9.1 with

$$h_n(f) = \frac{n}{\sqrt{2N}} \sum_{l \in \mathcal{N}} \theta_l^2.$$

If $h_n(f) = O(1),\ f \in \mathcal{F}$, then $\|f\| = o(1)$. In fact,

$$\|f\|^2 = \sum_{l \in \mathcal{L}} \theta_l^2 \le \sum_{l \in \mathcal{N}} \theta_l^2 + C^{-2} \sum_{c_l \ge C} c_l^2 \theta_l^2 \le \frac{\sqrt{2N}}{n} h_n(f) + C^{-2} = o(1).$$

These yield (9.2) for $f \in \mathcal{F}$ such that $h_n(f) = O(1)$. If $h_n(f) \to \infty$, then it follows from Chebyshev's inequality and the boundness of the variances that $P_{n,f}(U_n \ge H) \to 0$ for $H < ch_n(f),\ c \in (0,1)$. Hence, Theorem 1 (2) follows. $\square$

This completes the proof of Theorem 1.

References

[1] Abramovich, F., De Feis, I., Sapatinas, T. (2009). Optimal testing for additivity in multiple nonparametric regression. *Ann. Inst. Statist. Math.*, **61**, 691–714.

[2] Brown, L.D., Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, **24**, 2384–2398.

[3] Carter, A. (2006). A continuous Gaussian process approximation to a nonparametric regression in two dimensions. *Bernoulli*, **12**, 143–156.

[4] Gine, E.M. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.*, **3**, 1243–1266.

[5] Cohen, A. (2003). *Numerical Analysis of Wavelet Methods*. Elsevier, Amsterdam.

[6] Efromovich, S. (2000). On sharp adaptive estimation of multivariate curves. *Math. Methods Statist.*, **9**, 117–139.

[7] Ermakov, M.S. (1990). Minimax detection of a signal in a Gaussian white noise. *Theory Probab. Appl.*, **35**, 667–679.

[8] Ermakov, M.S. (2003). On asymptotic minimaxity of kernel-based tests. *ESAIM: Probab. Statist.*, **7**, 277–310.

[9] Guerre, E., Lavergne, P. (2002). Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory*, **18**, 1139–1171.

[10] Hall, P. (1984). Central limit theorem for integrated squared error for multivariate nonparametric density estimators. *J. Multivar. Anal.*, **14**, 1–16.

[11] Horowitz, J.L, Spokoiny, V.G. (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a non-parametric alternative. *Econometrica*, **69**, 599–631.

[12] Huang, J.Z. (1998). Projection estimation in multiple regression with application to functional ANOVA models. *Ann. Statist.*, **26**, 242–272.

[13] Ingster, Yu.I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problems Inform. Transmission*, **18**, 130–140.

[14] Ingster, Yu.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist.*, **2**, 85–114, 171–189, 249–268.

[15] Ingster, Yu.I. (1993). Minimax testing of the hypothesis of independence for ellipsoids in l_p. *Zapiski Naushn. Sem. POMI*, **207**, 77–97 (in Russian, translated in *J. Math. Sci.*, **81**, 2406–2420 (1996)).

[16] Ingster, Yu.I. (1994). Minimax hypotheses testing on a probability density for ellipsoids in l_p. *Theory Probab. Appl.*, **39**, 530–553.

[17] Ingster, Yu.I., Kutoyants, Yu.A. (2007). Nonparametric hypothesis testing for an intensity of Poisson process *Math. Methods Statist.*, **16**, 217–245.

[18] Ingster, Yu.I., Suslina I.A. (2002). *Nonparametric Goodness-of-Fit Testing under Gaussian Model.* Lectures Notes in Statistics. Vol. **169**, Springer-Verlag, New York.

[19] Ingster, Yu.I., Suslina I.A. (2005). On estimation and detection of smooth function of many variables. *Math. Methods Statist.*, **14**, 299–331.

[20] Ingster, Yu.I., Suslina I.A. (2007). Estimation and detection of high-variable function from Sloan-Woźniakowski space. *Math. Methods Statist.*, **16**, 318–353.

[21] Ingster, Yu.I., Suslina I.A. (2007). On estimation and detection of a function from tensor product spaces. *Zapiski Nauchn, Sem. POMI*, **351**, 180–218. (in Russian, translated in *J. Math. Sci.*, **152**, 897–920 (2008)).

[22] Krantz, S.G. (1992). *Function Theory of Several Complex Variables.* 2nd edition, Belmont, Wadsworth & Brooks/Cole.

[23] Kuo, F.Y., Sloan, J.H. (2005). Lifting the curse of dimensionality. *Notices of AMS*, **52**, 1320-1329.

[24] Levit, B., Stepanova, N. (2004). Efficient estimation of multivatiate analytic functions in cube-like domains. *Math. Methods Statist.*, **13**, 253–281.

[25] Lin, Y. (2000). Tensor product space ANOVA model. *Ann. Statist.*, **28**, 734–755.

[26] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, **24** , 2399–2430.

[27] Reiss, M. (2007). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.*, **36**, 1957–1982.

[28] Sloan, I.H., Woźniakowski, H. (1998). When are quazi-Monte Carlo algorithms efficient for high dimensional integrals? *J. Complexity*, **14**, 1–33.

[29] Spokoiny, V.G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.*, **24**, 2477–2498.

[30] Temlyakov, V.N. (1993). *Approximation of Periodic Functions.* Nova Science Publishers, New York.

[31] Walter, G.G. (1994). *Wavelets and Other Ortogonal Systems with Applications.* CRC Press, Boca Raton.

[32] Woźniakowski, H. (2006). Tractability of multivariate problems for weighted spaces of functions. In: *Appoximation and Probability, Banach Center Publications*, **72**, Institute of Mathematics, Polish Academy of Science, Warszawa.